2010

Computational and Experimental Investigations of Gas-Phase Thermodynamics of Non-Protein Amino Acids

Ashley R. Janiga
College of William & Mary - Arts & Sciences

Computational and Experimental Investigations of Gas-Phase Thermodynamics of Non-Protein Amino Acids

Ashley R. Janiga

Duryea, Pennsylvania

Bachelor of Science in Chemistry, King's College, 2008

A Thesis presented to the Graduate Faculty
of the College of William and Mary in Candidacy for the Degree of
Master of Science

Department of Chemistry

The College of William and Mary
August, 2010

APPROVAL PAGE

This Thesis is submitted in partial fulfillment of
the requirements for the degree of

Master of Science

Ashley R. Janiga

Ashley R. Janiga

Approved by the Committee, May, 2010

Committee Chair
Professor John C. Poutsma, Chemistry
The College of William and Mary

Professor Christopher J. Abelt, Chemistry
The College of William and Mary

Associate Professor Elizabeth J. Harbron, Chemistry
The College of William and Mary

ABSTRACT PAGE

The present work is part of a continuing effort from our laboratory to provide additional insights into the thermochemical properties of non-protein amino acids by means of intrinsic molecular properties such as gas-phase acidities, gas-phase basicities, and proton affinities.

The proton affinities of cysteine and serine analogs have been measured using Cooks' extended kinetic method in a quadrupole ion trap instrument. Experimental values of 919.8 ± 12.8 kJ/mol and 942.5 ± 18.0 kJ/mol have been determined for the 298K proton affinities of homocysteine and homoserine, respectively. High level computational calculations using B3LYP methods at a variety of basis sets were carried out in order to give theoretical predictions for the 298K proton affinities and gas-phase acidities of the cysteine and serine analogs. Recommended values for the gas-phase acidities and proton affinities for homocysteine and homo homocysteine were calculated to be 1397.2 and 1408.5 kJ/mol and 931.7 and 945.2 kJ/mol, respectively. Gas-phase acidities and proton affinities of homoserine and homo homoserine were calculated to be 1393.6 and 1383.4 kJ/mol and 942.1 and 975.0 kJ/mol, respectively.

The proton affinities of the amide carbonyl oxygen and the amide nitrogen of proline homologs, acetylproline, acetylpipecolic acid, acetylazetidine-2-carboxylic acid, and acetylNMA, have been determined using hybrid density functional theory calculations. These computational results show that the amide carbonyl oxygen has a larger proton affinity than that of the amide nitrogen.

This thesis is dedicated to my parents, without whom I wouldn't be here (in more ways than one). Their unwavering love and support is one of the driving forces in my life, and I can never repay them.

Table of Contents

Acknowledgements

This thesis would not have been possible without the love and support of a number of incredibly special people.

First of all, I wish to express my love and gratitude to my beloved family, for their understanding and endless love, through the duration of my studies. To my parents for allowing me to realize my own potential. All the support they have provided me over the years was the greatest gift anyone has ever given me. I must acknowledge my boyfriend, Jonathan, without whose love and encouragement, I would not have finished this thesis.

I would like to express my sincere gratitude to my advisor, Dr. J.C. Poutsma who was abundantly helpful and offered invaluable assistance, support, and guidance throughout this journey. Deepest gratitude is also due to the other members of the Ionlab who created such a wonderful atmosphere in the lab.

I also need to further thank all my friends, particularly Gina, Dan, and Mike. You have always been there for me, whether we live in the same state or hundreds of miles away. No matter how far away I am I know that you will always be my friends.

Finally, I would like to acknowledge my committee members, Dr. Abelt and Dr. Harbron, who graciously agreed to serve on my committee. Although you are busy in other endeavors, I thank you for your valuable time and insights.

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Protein Composition and Structure

Proteins, like other biological macromolecules, are crucial components of organisms and participate in essentially all biological processes. They function as catalysts, provide mechanical support and immune protection, transport and store other molecules, generate movement, transmit nerve impulses, and control growth and differentiation.[1]

Proteins are linear polymers composed of amino acids that are linked end to end. The amino acids are joined together by peptide bonds between the carbonyl and amino groups of adjacent amino acid residues. Most proteins spontaneously fold into unique 3-dimensional structures that are determined by the amino acid sequence. These structures are shaped by hydrogen bonding, disulfide bridges, Van der Waals forces, and electrostatic interactions. Pace[2] and coworkers have suggested that the intrinsic properties of the amino acid side chain are a leading contributor in the direction of protein folding. The chemistry of the side chains present in the binding pocket also defines the proteins catalytic active sites. Further investigations into the intrinsic acid/base properties of amino acids are essential in the comprehension and prediction of protein folding behavior and enzyme active site reactivity.

1.2 Mass Spectrometry

1.2.1 – History and Biological Applications

Mass spectrometry is a powerful analytical technique that is used to elucidate the structure and chemical properties of molecules, identify unknown compounds, and quantify known compounds. The first step in the mass spectrometric analysis of compounds is the production of gas phase ions. Early mass spectrometers required a gaseous sample, but due to recent advancements, samples in solution or embedded in a solid matrix are now commonly analyzed. Upon formation of the gas phase ions, the sample enters the vacuum chamber through an inlet. It is essential that all mass spectrometers function under high vacuum in order to allow ions to reach the detector without undergoing collisions with the other gaseous molecules. From the vacuum chamber, the gas phase ions enter the mass analyzer where they are separated according to their mass-to-charge ratio. The ions are then detected and transformed into a usable signal by a detector. A mass spectrum of the molecule is thus produced.

The history of mass spectrometry dates back more than one hundred years beginning with the studies of Sir Joseph John Thomson. Thomson's investigations led to the discovery of the electron in the late 1800s and the construction of the first mass spectrometer in the early 1900s. Later, Thomson's student, Francis W. Aston, designed a mass spectrometer with increased resolution. Simultaneously, A.J. Dempster fabricated a mass spectrometer which also had an improved resolving power and developed the first electron ionization source.[3-7]

In Dempster's mass spectrograph (Figure 1.1), molecules are first ionized by a hot filament (G) and then accelerated towards the entrance slit (S_1). The ions then

follow a hemispherical trajectory in which the radius of the trajectory is defined by three slits (S_1, S_2, and S_3). Only ions with this trajectory are then detected by the detector (E).[6]

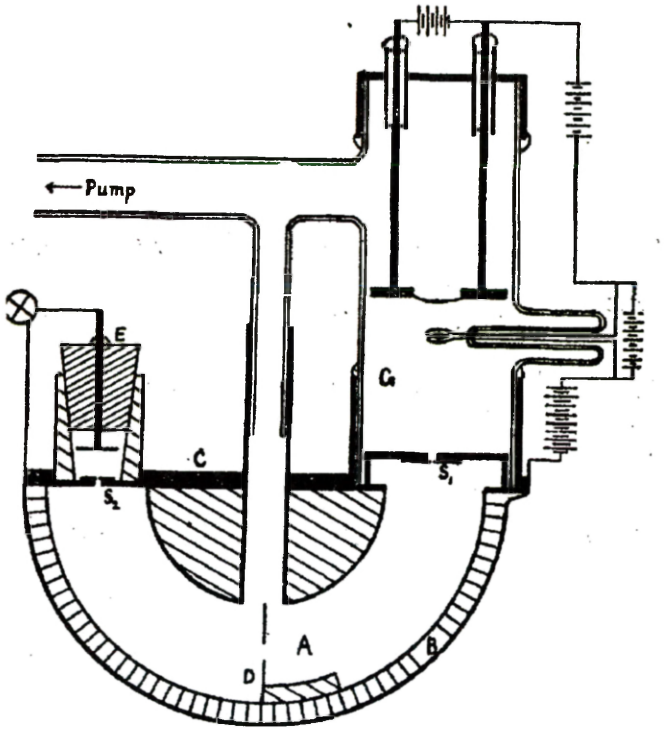


Figure 1.1 – Dempster's Mass Spectrograph[6]

In the early 1950's, Wolfgang Paul recognized that the use of magnetic fields could be eliminated if an alternating quadrupolar electric field was used.[8] This quadrupole ion trap was developed in parallel with the quadrupole mass analyzer. A schematic of this instrument can be seen in Figure 1.2.

Quadrupoles can also be placed in tandem to enable them to perform fragmentation studies. The most common set-up is the triple quadrupole mass spectrometer (QqQ). Quadrupole mass spectrometers are symbolized by upper case

Q and RF only quadrupoles with a lower case q. A collision gas can be introduced into the second quad in order to induce fragmentation.[9]

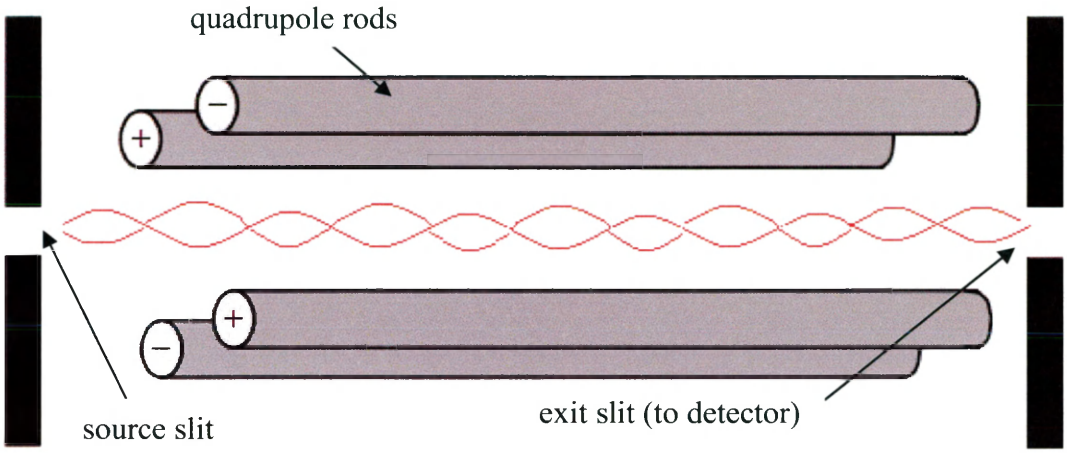


Figure 1.2 – Schematic of a Quadrupole Mass Analyzer

The quadrupole ion trap is used to isolate and characterize molecules utilizing the individual mass to charge ratios. A schematic of an ion trap can be viewed in Figure 3.1 in Chapter 3. Recently, the operation of the quadrupole ion trap in the mass-selective instability mode has become increasingly valuable in chemical analysis. In this method, an r.f. voltage is applied to the ring electrode, while the two end cap electrodes are held at the ground potential. All ions in which the mass to charge ratios exceed a minimum cut-off value are trapped. To record a mass spectrum, the r.f. voltage is increased with respect to time so that ions of successively greater *m/z* ratios develop unstable trajectories.[9] These ions are ejected through perforations in an end cap and are then detected at an electron multiplier.

Ions can be generated within the cavity of the ion trap using electron impact, positive chemical ionization or negative chemical ionization. In addition, modern ionization methods such as laser desorption, secondary ion mass spectrometry, and electrospray ionization have been used in conjunction with an ion trap allowing for the analysis of previously unattainable molecules (i.e. involatile samples).

In addition to providing means to determine accurate masses and sequence information, electrospray ionization (ESI) mass spectrometry provides opportunities to study the gas-phase characteristics of biomolecules. Ionization is the process in which a gas-phase ion is generated from a liquid chemical species. This is achieved by spraying a sample solution from a heated capillary into a strong electric field under atmospheric pressure.[9]

The extent to which a protein or peptide is protonated or deprotonated in the gas phase can provide insights into important aspects of the molecular structure, stability, and Brønstead acid/base reactivity. Generally, protons reside on the side chains of basic amino acids and the N-terminal amino acid. Therefore, the number and accessibility of the amino acids in the protein can be found from an ESI spectrum.

Proton transfer reactions have been used to elucidate thermodynamic properties such as apparent proton affinity (PA), gas-phase basicity (GB), and gas-phase acidity (GA) for singly and multiply charged molecules.[10]

In our lab, proton bound dimers are isolated in the quadrupole ion trap's cavity and subjected to collision induced dissociation. The product fragments are then detected in proportion to the affinity of each monomer to the proton.

1.2.2 – Methods of Determining Proton Affinities

The earliest gas-phase estimates of proton affinities have been determined from measurements in which the direction of exothermic proton transfer between base pairs is established. By reacting the species B_iH^+ with a variety of bases with known proton affinities, it is possible to bracket $PA(B_i)$.[11]

Another technique employed in the determination of proton affinities is the equilibrium method. In this method, proton affinities are determined from measurements of equilibrium constants for gas phase proton transfer between two bases B_1 and B_2. This can be accomplished by measuring the rate constants for the forward and reverse reactions or by measuring the equilibrium constant directly. These equilibrium measurements require that the equilibrium ion signals be measured on an absolute basis and the concentrations of the neutral bases be measured accurately.

From the equilibrium constant, ΔG° can be determined. Absolute basicities can be ascertained by creating ladders of relative gas-phase basicities which include reference compounds of known basicity. The difference in proton affinities can then be established if the equilibrium constant is measured as a function of temperature.[11]

Although the vast majority of proton affinities have been determined by the equilibrium method, it is not useful for involatile molecules such as amino acids and peptides due to the inability to measure accurate concentrations of the neutral bases.

The kinetic method approach, however, can be utilized in the determination of the proton affinities of biomolecules. This method involves the study of the

competitive dissociation of a weakly bonded heterodimer that can give rise to two possible protonated species.

In this present work, the application of the kinetic method in the determination of the proton affinities of non-protein amino acids is described.

1.3 Non-protein Amino Acids

Non-protein amino acids (NPAAs) are naturally occurring compounds found mainly in the plant kingdom.[12] Although many NPAAs are structurally analogous to the twenty protein amino acids, they are not used by humans in the formation of peptides and proteins. Most NPAAs are the byproducts of plant metabolism and serve a variety of functions ranging from predatory defense to nitrogen fixation mechanisms. Due to the similarity in structure to one or more of the twenty protein amino acids (PAAs), NPAAs can compete with them in a variety of biochemical pathways. In fact, many NPAAs are toxic to mammals.[13]

In order to facilitate the understanding of the composition of amino acids and their intrinsic chemistry, gas-phase thermochemical properties of NPAAs have been studied using the extended kinetic method in an electrospray ionization-quadrupole ion trap instrument.

Recent studies by Poutsma and coworkers have investigated the intrinsic properties of proline and lysine homologs. Proline analogs, L-azetidine-2-carboxylic acid, L-proline, and L-pipecolic acid have been studied in order to determine the effect of ring size on the proton affinity of these amino acid analogs. Poutsma and

coworkers have reported values of 933.0, 941.0, and 943.9 kJ/mol for the 4-, 5-, and 6-membered ring proline analogs.[14]

The proton affinities of lysine and its three homologs ornithine, 2,4-diaminobutanoic acid, and 2,3-diaminopropanoic acid have also been reported. A value of 1004.2 kJ/mol is recommended for the proton affinity of lysine and values of 1001.0, 975.8, and 950.2 kJ/mol have been determined for ornithine, 2,4-diaminobutanoic acid, and 2,3-diaminopropanoic acid respectively.[15]

1.4 Proteomics

In the past decade, a technological revolution has taken place culminating in the sequencing of the human genome and increasing interest in proteomics, the study of proteins. The term proteomics was coined by Williams and coworkers in the early 1990s. The word is derived from PROTEins expressed by a genOME and it refers to all the proteins produced by an organism.[16]

Cells express genes, which encode proteins with functions unique to specific cells in addition to the proteins that carry out generic functions. The affinity, function, cellular abundance, and transport of proteins are all influenced by a wide variety of reversible and irreversible chemical reactions. Approximately 200 different post-translational modifications influence these chemical reactions and make up an organisms proteome. Due to these modifications, analysis of the proteome is very complex.

Mass spectrometry-based proteomics allows for the identification and precise quantification of thousands of proteins from complex samples. By sequencing and

identifying smaller pieces of the protein, the identity and the amino acid composition can be determined. Each amino acid, except leucine and isoleucine, has a unique mass, making identification relatively easy.

Proton-transfer processes are elementary reactions of fundamental importance in chemistry and biology. Knowledge of the proton affinity of a chemical species is highly useful in modeling and analyzing the thermodynamic properties of a given mechanism. The work in this thesis describes the computational and experimental methods utilized in the determination of the proton affinities of cysteine and serine homologs and acetylated proline analogs. Experimental values for the proton affinities of homocysteine and homoserine were determined using mass spectrometry and Cooks' extended kinetic method. In addition, three computational studies are described. One in which the proton affinities and gas-phase acidities of homocysteine (2-amino-4-sulfanylbutanoic acid) and homoserine (2-amino-4-hydroxybutanoic acid) were predicted for comparison to experimental values. The proton affinities and gas-phase acidities of homo homocysteine (2-amino-5-sulfanylpentanoic acid) and homo homoserine (2-amino-5-hydroxypentanoic acid) were also calculated in order to further the understanding of the affects of side chain length on thermochemical properties of a molecule. The final computational study was performed on acetylated proline analogs in an effort to understand the governing factors of the "proline effect".

Chapter 2 – Kinetic and Computational Methods

2.1 Kinetic Method: Proton Affinity Determinations

Thermochemical properties of organic molecules have been studied quantitatively for more than 3 decades and are imperative in the understanding of the chemical behavior of these molecules. Gas-phase analysis of these chemical properties is highly effective due to the absence of a solution matrix. This absence allows for the analysis of the intrinsic properties of just the analyte.

Gas-phase proton transfer (equation 1) is essential in understanding the intrinsic effects on the acidity and basicity of organic molecules.

$$AH^+ + B_i \longrightarrow A + B_iH^+ \quad (1)$$

The relationship between the enthalpy of formation of MH^+ and its neutral counterpart, M, is defined in terms of a proton affinity, the negative change in enthalpy (ΔH) associated with protonation. This can be seen in equation (2).

$$M + H^+ \longrightarrow MH^+ \qquad PA \equiv -\Delta H_{reaction} \quad (2)$$

The Gibbs energy change, shown in equation (3), associated with the protonation reaction above, is called the gas-phase basicity, GB, of a molecule.

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

Given that the gas-phase basicity can be expressed in terms of the temperature dependent entropy change and the proton affinity, equation (3) can be rewritten as shown in equation (4).

$$GB = PA + T\Delta S \quad (4)$$

The kinetic method is an approximate, semiquantitative method of acquiring thermochemical parameters by mass spectrometry. In this method, the proton bound heterodimer, $A - H^+ - B_i$, is the species of interest. Dissociation of this heterodimer may give rise to two protonated species, AH^+ and B_iH^+, via the competitive dissociation channels k_A and k_{Bi}, assuming no reverse activation energy and only structurally similar analytes and reference bases are involved. The dissociation channels for proton affinities are given by equation (5):

$$AH^+ + B_i \xleftarrow{k} [A - H - B_i] \xrightarrow{k_i} A + B_iH^+ \quad (5)$$

In the above equation, k represents the rate coefficient for the formation of AH^+ with unknown gas-phase thermochemistry, while k_i refers to the rate coefficient for one of a series of structurally-similar reference bases having known values for the proton affinity, gas-phase basicity (GB), and protonation entropy (ΔS).

Similarly the gas-phase acidities can be determined by decomposition of a negatively charged heterodimer as shown in equation (6).

$$A^- + B_iH \xleftarrow{k} [A - H - B_i]^- \xrightarrow{k_i} AH + B_i^- \quad (6)$$

The logarithm of rate coefficients, from equation (5), can be expressed using unimolecular reaction theory because the competitive dissociations occur from the common heterodimer, $A\text{-}H^+\text{-}B_i$. Unimolecular reaction theory relates the logarithm of the ratio of rate coefficients to the difference of the natural logarithm of transition state partition functions (Q^*) for each protonated molecule and the critical energy of the unimolecular dissociation ($\Delta\varepsilon_0$) as shown in equation (7).[17,18]

$$\ln\frac{k_i}{k} = \ln\left(\frac{[B_iH^+]}{[AH^+]}\right) \approx \ln\left(\frac{Q_i^*}{Q^*}\right) + \frac{\Delta\varepsilon_0}{RT_{eff}} \quad (7)$$

Q_i^* and Q^* refer to the partition functions of the transition states of the two dissociation channels and the effective temperature, T_{eff}, represents the characteristic temperature of activated complexes that undergo competitive fragmentation. The term $\Delta\varepsilon_0$ is the difference in critical energies for the two dissociation channels. Once again, assuming no reverse activation energy, the $\Delta\varepsilon_0$ term can be replaced by the difference in proton affinities of the reference base and analyte. As can be seen in equation (8).

$$\ln\frac{k_i}{k} = \ln\left(\frac{[B_iH^+]}{[AH^+]}\right) \approx \ln\left(\frac{Q_i^*}{Q^*}\right) + \frac{PA(B_i) - PA(A)}{RT_{eff}} \quad (8)$$

In the simplest form of the kinetic method, the assumption is made that the term $\ln\left(\frac{Q_i^*}{Q^*}\right)$ is equivalent to the difference in protonation entropy between the two competitive dissociation channels.[19,20] This can be seen in equation (9),

$$\ln\frac{k_i}{k} = \ln\left(\frac{[B_iH^+]}{[AH^+]}\right) \approx \frac{PA(B_i) - PA(A)}{RT_{eff}} + \frac{\Delta(\Delta S)}{R} \quad (9)$$

where $\Delta(\Delta S)$ is defined as the reaction entropy difference between the two fragmentation channels. With the assumption that only structurally similar analytes and reference bases are involved, the entropic term can be canceled leading to the standard version of the kinetic method shown in equation (10),

$$\ln\frac{k_i}{k} = \ln\left(\frac{[B_iH^+]}{[AH^+]}\right) \approx \frac{PA(B_i) - PA(A)}{RT_{eff}} \quad (10)$$

in which $\ln \frac{B_i H^+}{AH^+}$ plotted against PA(B_i) results in a straight line with a slope of $\frac{1}{RT_{eff}}$ and an intercept of $\frac{-PA(A)}{RT_{eff}}$. The proton affinity of the unknown can be derived simply as the ratio of the negative value of the intercept to the slope.[19,20]

Fenselau[21] and Wesdemiotis[22], in attempts to apply the kinetic method to biomolecules, realized that obtaining a series of reference bases with a similar structure to the analyte would be difficult. It was found that while the entropic term, $\frac{\Delta(\Delta S)}{R}$, was not negligible due to the negation of the second assumption, it will likely remain constant if the reference bases were structurally similar among themselves. Thus a plot of $\ln \frac{k_i}{k}$ versus PA(B_i) should yield a straight line with a slope of $\frac{1}{RT_{eff}}$ and an intercept of the apparent gas-phase basicity, $GB^{app}(A)$. These quantities are related through equation (11).

$$\frac{GB^{app}(A)}{RT_{eff}} \approx \frac{PA(A)}{RT_{eff}} - \frac{\Delta(\Delta S)}{R} \qquad (11)$$

A second plot is constructed by plotting the negative of the intercepts versus the slopes obtained from a series of experiments performed at varying collision energies in order to obtain PA(A) and $\Delta(\Delta S)$. Armentrout and coworkers have pointed out that a statistical analysis of the second plot demonstrates that the excellent linear correlations ($r^2 = 0.99999$) are artifacts of the choice made in plotting the data. In order to remove the correlation between the slope and intercept of the linear regression analysis, they have developed a statistical procedure to remove this covariance. Instead of plotting $\ln \frac{k_i}{k}$ versus PA(B_i), the average proton affinity of the

reference bases, $PA(B_i^{avg})$, is subtracted from the proton affinities of the reference bases and analyte in equation (9) to give the modified equation (12).

$$\ln\frac{k_i}{k} \approx \frac{PA(B_i) - PA\left(B_i^{avg}\right)}{RT_{eff}} - \left[\frac{PA(A) - PA\left(B_i^{avg}\right)}{RT_{eff}} - \frac{\Delta(\Delta S)}{R}\right] \qquad (12)$$

A plot of $\ln\frac{k_i}{k}$ versus $\Delta PA(B_i)$ - $\Delta PA(B_i^{avg})$ removes the covariance between the slope, $\frac{1}{RT_{eff}}$, and the intercept, $-\left[\frac{PA(A)-PA\left(B_i^{avg}\right)}{RT_{eff}+\Delta-\left[\Delta S/R\right]}\right]$, that can be seen in Armentrout and coworkers initial plot.

A temperature corrected kinetic method (TCKM) has recently been proposed.[23] A triple plotting procedure would be used to obtain improved relative thermochemical measurements. Although this method gives excellent data, it incorrectly represents the errors through extreme averaging. In cases where reference bases are not structurally similar, the TCKM procedure will not yield a straight line in the plot of $\ln\frac{k_i}{k}$ versus $PA(B_i)$ due to the varying $\Delta(\Delta S)$ term.

The entropy-corrected kinetic method (equation 13) allows for the limitations of the kinetic method to be overcome. By incorporating the entropic term, an entropy-corrected logarithm of the branching ratio term $\ln\left(\frac{k_i}{k}\right) - \frac{\Delta S(B_i)}{R}$ is obtained. To remove the correlation between the slope and intercept of the linear regression analysis, Cooks' and coworkers have adopted Armentrout's improved double plotting method.[24]

$$\ln\left(\frac{k_i}{k}\right) - \frac{\Delta S(B_i)}{R} \approx \frac{PA(B_i) - PA\left(B_i^{avg}\right)}{RT_{eff}} - \left[\frac{PA(A) - PA\left(B_i^{avg}\right)}{RT_{eff}} + \frac{\Delta S(A)}{R}\right] \qquad (13)$$

This extended version of the entropy-corrected kinetic method not only takes into account the entropy effects but can also provide relative reaction entropies in the instances where the reference bases are similar in structure to each other but are dissimilar to the analyte. Cooks' entropy-corrected extended kinetic method equation will yield a straight line with a slope of $\frac{1}{\mathrm{RT_{eff}}}$ and an intercept given by $-\left[\frac{\mathrm{PA(A)}-\mathrm{PA}(\mathrm{B}_i^{\mathrm{avg}})}{\mathrm{RT_{eff}}}+\frac{\Delta \mathrm{S(A)}}{R}\right]$ when the entropy-corrected term, $\ln\left(\frac{k_i}{k}\right)-\frac{\Delta \mathrm{S}(\mathrm{B}_i)}{R}$, is plotted against $\mathrm{PA}(\mathrm{B}_i)-\mathrm{PA}(\mathrm{B}_i^{\mathrm{avg}})$.

Ervin and Armentrout[12] have recently developed a new approach to fitting kinetic method plots involving an orthogonal distance regression (ODR) approach.[24] In this method, a regression algorithm is used to force a single isothermal point in the first kinetic method plot. The isothermal point exists when the equilibrium constant $\mathrm{K}=\frac{k_{\mathrm{B}_i}}{k_A}\approx\frac{\mathrm{B}_i}{\mathrm{A}}$ is the same for all effective temperatures. Because the proton affinity as well as the entropy is obtained from the isothermal point, the error in this point is the experimental error.

In the ODR approach, the numbers of best fit lines, n, at m different energies are forced to cross at a common point. A Monte Carlo simulation, which better allows for the estimation of the uncertainties in the experimental ion ratios and the proton affinity of each reference base, is executed by generation of random noise within user-defined ranges.[26,27] The final ODR analysis gives values for entropy and enthalpy that coincide with the entropy-corrected extended kinetic method provided that a sharp isothermal point is present.

2.2 Computational Chemistry

Computational chemistry is used in a number of different ways in academia and industry. It is used to model a molecular system prior to synthesizing that molecule in the laboratory and also used to understand properties of a molecule that can be more easily obtained computationally than experimentally. One such property is molecular bonding. Thus, computational modeling is now being used to gain additional understanding of compounds that are being examined in the laboratory.

Computational chemistry is built on the fundamental principles of quantum mechanics, the correct mathematical description of the behavior of microscopic particles such as electrons and nuclei. In theory, quantum mechanics is able to predict a property of an individual molecule or atom exactly. In reality, the quantum mechanical equations have only been solved exactly for one electron systems.

In quantum mechanics, the Schrödinger equation describes how the quantum state of a physical system changes with respect to time. The time independent version of the Schrödinger equation is shown in equation (14)

$$\hat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{14}$$

where $\hat{H}$ and Ψ satisfy:

$$\hat{H}\Psi = E\Psi \tag{15}$$

In the above equation, $\hat{H}$ is the Hamiltonian operator, Ψ is the wave function, and E is the energy. The wave function, Ψ, is a function of the electronic and nuclear position and is a probabilistic description of electron behavior.[29]

Due to the complexity of the mathematical equation, computers are used in conjunction with simplifications and parameterization to solve the equations. In the currently available software, the full Hamiltonian is rarely used. Instead, the Born-Oppenheimer approximation is utilized for simplification purposes. In the Born-Oppenheimer approximation, the motions of the electrons are separated from the motions of the much heavier nuclei. As a result, nuclear positions can be used as parameters to determine the motion of the electrons. In this approach, the atoms are arranged and the electrons are allowed to distribute themselves in the available orbitals in order to maximize the Coulombic attraction between the electrons and the nuclei while minimizing the Coulombic repulsion between electrons. The nuclei are then systematically moved around until the optimal geometry is found. For each electronic calculation, a basis set must be supplied. The basis set is an arrangement of orbitals that the electrons are allowed to occupy. The choice for the mathematical representation is arbitrary as long as the chosen functions obey certain rules.

The computational methods for use in the study of molecules can be divided into three categories: *ab initio*, semiempirical, and force field models.

The term *ab initio* is Latin for "from the beginning." This name is given to computations which are solely based on the laws of quantum mechanics, the masses and charges of electrons and atomic nuclei, and the values of fundamental physical constants. *Ab initio* methods are approximate quantum mechanical calculations which include Hartree-Fock (HF) theory, configuration interaction (CI) theory, perturbation theory (PT), and density functional theory (DFT). *Ab initio* molecular orbital computations can provide accurate quantitative predictions of chemical

properties for a wide range of molecular systems. However, they are extremely demanding on computer resources, especially for large molecular systems.[29]

Density functional theory has become very popular for quantum mechanical calculations in recent years. Unlike most *ab initio* methods, DFT is less computer intensive while maintaining similar accuracy. The principle behind the density functional theory is that the energy of a molecule can be determined from the electron density instead of a wave function. In this method, the electron density is expressed as a linear combination of basis functions similar in form to that of the Hartree-Fock orbitals. A determinant is then formed from these functions. It is the electron density from the determinant that is used to compute the energy. This process is essential because Fermion systems can only have electron densities that arise from an antisymmetric wave function.

Hybrid functionals are a class of approximations that combine functionals from other methods with pieces of a Hartree-Fock calculation, usually the exchange integrals. The parameters determining the weight of each individual functional are typically specified by fitting the functional's predictions to accurately calculated thermochemical or experimental data. One hybrid functional most often chosen for theoretical calculations is B3LYP. In this functional, the exchange energy comes from the Becke exchange functional and is combined with the exact energy from the Hartree-Fock theory.[29] In association with the component exchange and Lee, Yang and Parr correlation functionals, three parameters define the hybrid functional, specifying how much of the exact exchange is mixed in.[29]

Semiempirical methods increase the speed of computation by using approximations of *ab initio* techniques. They are set up with the same general structure as Hartree-Fock calculations in that they have a wave function and a Hamiltonian operator. Within this framework, however, certain pieces of information are approximated or intentionally omitted in order to simplify the calculation. In most cases, it is the core electrons that are not included in the calculation. To further simplify this method, a minimal basis set is used. In order to correct for these inaccuracies, parameters are used to estimate the omitted values by fitting the results to experimental data or *ab initio* calculations.

The Austin Model 1 (AM1) method has widely been used for the modeling of organic compounds. Depending on the nature of the system and information desired, AM1 will often give accurate results obtainable for organic molecules with semiempirical methods. It is with this method that our computational studies will begin.[29,30]

Semiemperical calculations can be used for large organic molecules; however, these calculations are too computationally intensive for most biomolecules. Nevertheless, it is still possible to model the molecules behavior circumventing quantum mechanics.

Molecular mechanics uses an energy expression that consists of a simple algebraic equation for the energy of a compound. In this technique, atoms are treated as spheres whose mass is dependent on the element. Chemical bonds are considered springs whose stiffness depends on not only the elements bound together but also the bond stability. Changes in bond angles and dihedral angles are also modeled by

various springs. Each of these springs will have constants associated with them. A wave function or the total electron density is not needed as the constants in this equation are obtained either from spectroscopic data or *ab initio* calculations. These parameters are calculated using force fields, a set of equations with their associated constants. Different force fields have been developed for different molecular types.[30]

To assist in the investigation of gas-phase thermochemical properties of non-protein amino acids, semiempirical and hybrid density functional theory calculations have been performed.

Chapter 3 – Experimental and Computational Procedures

3.1 Experimental Procedure

All experiments were performed in a commercial ion trap mass spectrometer (LCQ-DECA, Finnigan MAT, San Jose, CA) equipped with an external electrospray ionization source. Solutions of the analyte compounds and reference bases were prepared with an approximate concentration of 5×10^{-4} M. For positive ion solutions, a 49.5:49.5 methanol:water solvent was used. To facilitate the formation of positive proton-bound dimer complexes, 1% acetic acid was added. In order to obtain the proton affinity of the molecule of interest; the mass spectrometer was operated in positive ion mode.

Solutions were injected into an electrospray ionization source via a 500 µL Hamilton Gastight® syringe by an automated syringe pump at a flow rate of 900 µL/hr. Electrospray and ion-focusing conditions were varied to maximize the ion count for the proton-bound heterodimer. The proton-bound dimer ions were isolated at $q_z = 0.250$ with a mass-width adjusted to maximize ion signal while still maintaining isolation. A Finnigan LCQ™ DECA quadrupole ion trap mass spectrometer was used for the data collection. The heated capillary was maintained at a temperature of 125 °C and the voltage of the source and offset of the focusing lenses were set by the LCQ Tune™ software that served as the user interface to the mass spectrometer. Using this interface, the mass to charge ratio of the target heterodimer was input and the voltage and lenses were adjusted to provide optimal intensities. A 5 kV charge is applied to the electrospray needle to produce charged

micro droplets. These droplets, propelled by the nitrogen sheath gas (20 arbitrary units), pass through a heated capillary where, due to solvent evaporation and Coulombic explosion, they are converted into bare ions. R.f. voltages then attract them into the focusing octopoles and into the quadrupole ion trap. Within the trap, molecules with specific m/z can be isolated and collision-induced dissociation (CID) performed. In collision-induced dissociation experiments, molecular ions are fragmented in the gas-phase. The molecular ions are accelerated by an electrical potential to high kinetic energy in the vacuum of the mass spectrometer. These ions are then allowed to collide with neutral gas molecules, helium in this case. In this collision, some of the kinetic energy is converted into internal energy resulting in the breaking of bonds and the fragmentation of the molecular ions. These fragment ions are then analyzed by the mass spectrometer. A simplified diagram of this instrument can be seen below in Figure 3.1:

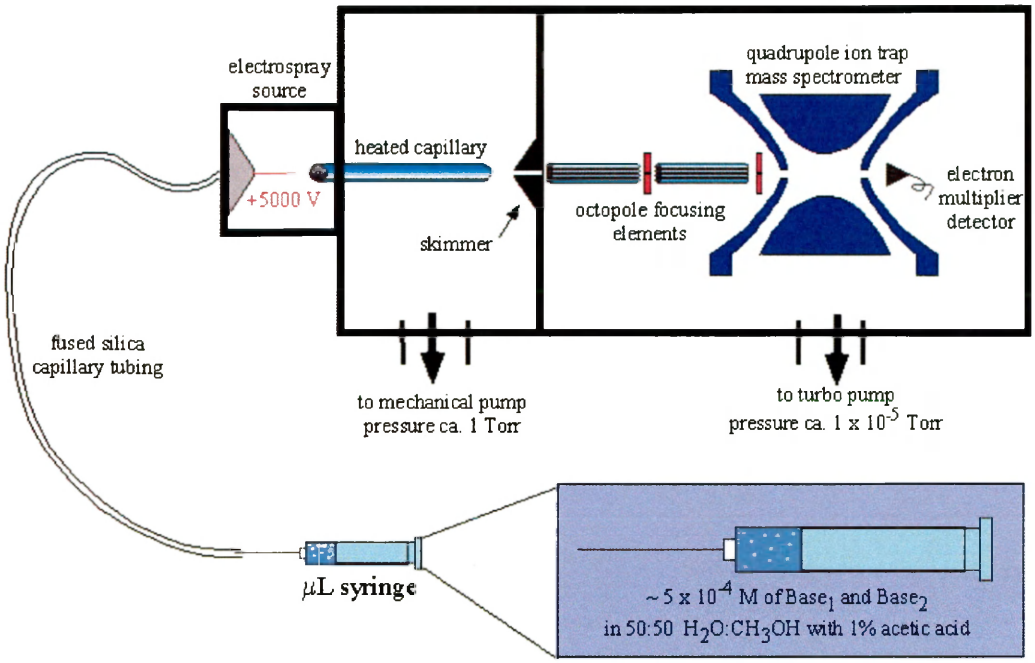


Figure 3.1 – Simplified Schematic of ESI Ion Trap Mass Spectrometry

MS spectra were evaluated to verify that the heterodimer, $[A\text{-}H^{+}\text{-}B_i]$ species, existed. Other peaks typically observed in the spectra included $A\text{-}H^{+}$, $A\text{-}H^{+}\text{-}A$, $B_i\text{-}H^{+}$, and $B_i\text{-}H^{+}\text{-}B_i$. Using the MS^n capabilities of the ion trap, the heterodimer peak was isolated and then fragmented using CID with helium gas at variable activation amplitudes. The variable activation amplitudes correspond to a percentage (i.e. 0-100%) of the maximal 5 kV applied voltage in the laboratory frame. The ion counts of the resulting products, A and B, were recorded at every 2% increase in activation amplitude up to 100%. The intensities of these peaks are dependent on the portion of each base population that remained protonated after the collision. The peak with the larger intensity has a higher affinity for the proton and therefore retains the proton more frequently. Figures 3.2-3.4 illustrate this process. The ratios of B/A, for each

of the 51 data points, are imported into Microsoft® Excel to be analyzed using Cooks' extended kinetic method analysis as described in equation (2). The final ion ratios are averages of at least three scans obtained on several different days.

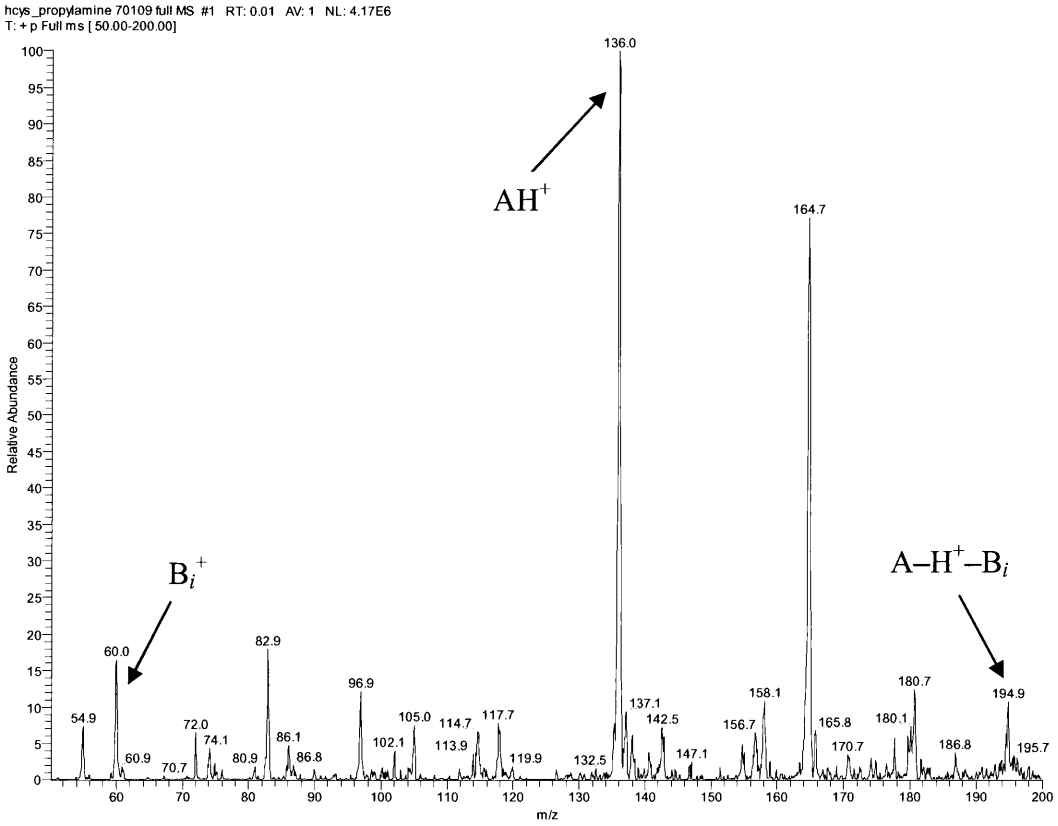


Figure 3.2 – Sample Mass Spectrum

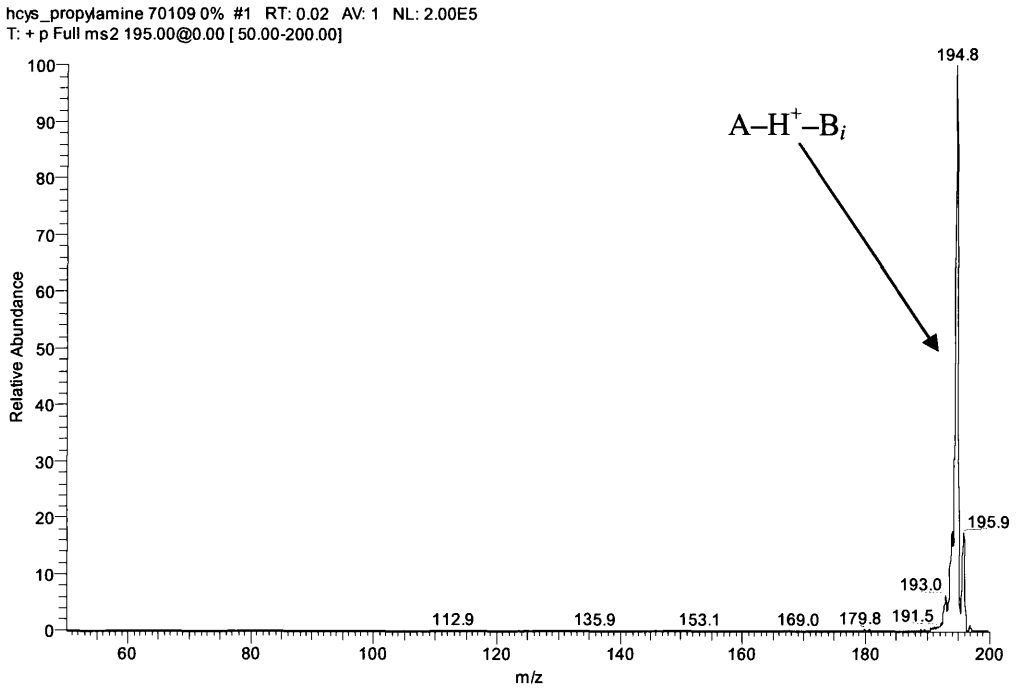


Figure 3.3 – Sample Isolation Spectrum

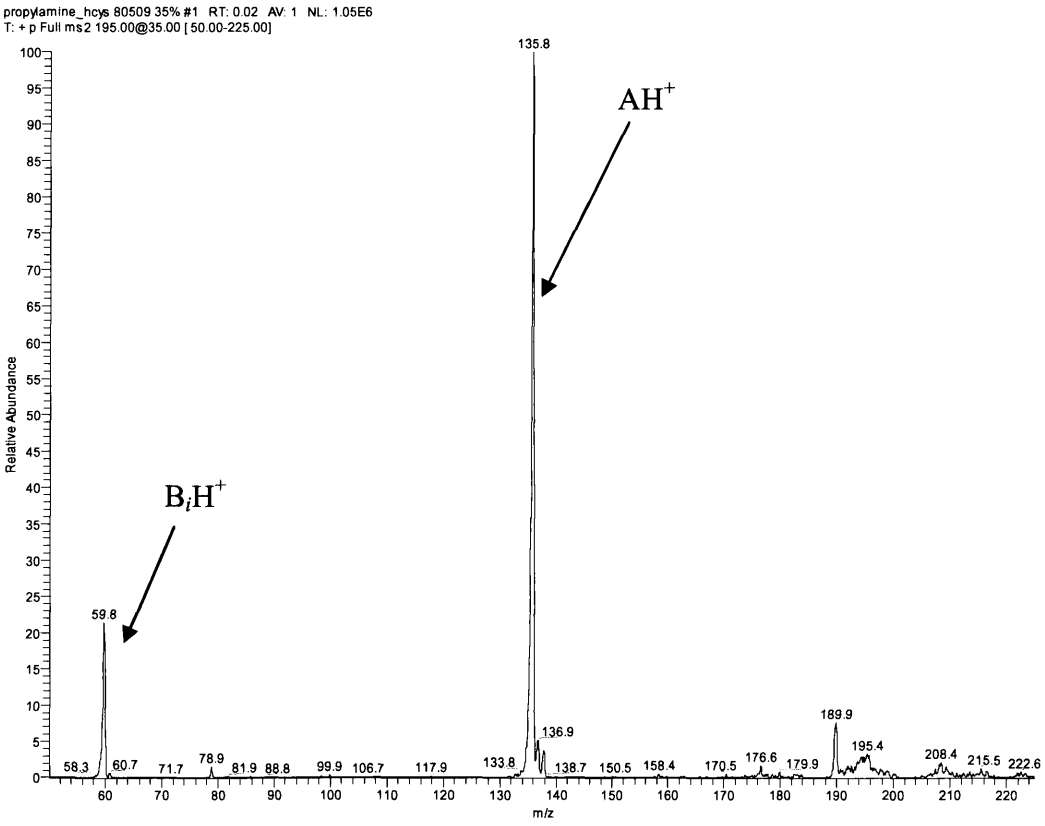


Figure 3.4 – Sample Fragmentation Spectrum

3.2 Computational Procedure

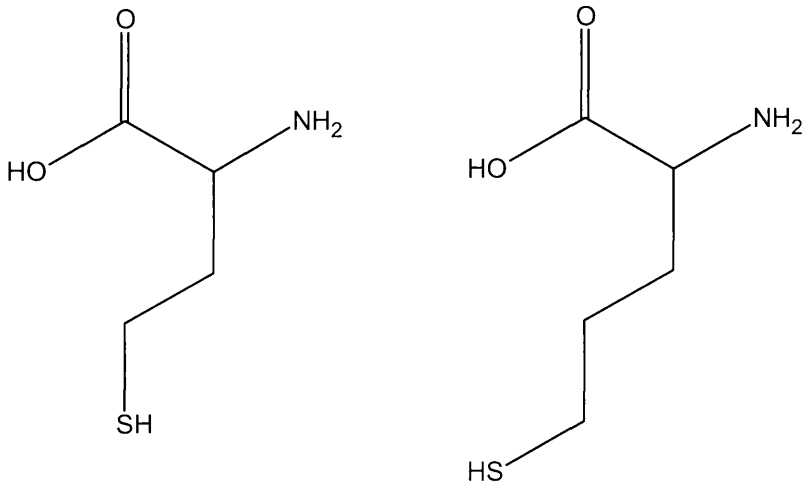
The theoretical analysis of each homo amino acid began with the PCModel® software. This software generates energy-minimum structural conformations using the GMMX conformational search algorithm. The GMMX routine generates random conformations through rotations about all single bonds of a molecule. The MMX force field is used to evaluate the energy of each of the random conformations. For this study, all conformations within 8 kcal/mol of the lowest energy conformer were used as starting structures for *ab initio* Hartree-Fock and hybrid B3LYP density functional theory calculations to find the lowest energy structures using Gaussian98®.[31-33] The structures were subjected to geometry optimizations starting at AM-1. The complexity of the basis sets was increased to RHF/3-21G, B3LYP/3-21G, and B3LYP/6-31+G* levels of theory. Vibrational frequencies and optimized geometries were calculated at the B3LYP/6-31+G* level of theory. Thermal corrections to enthalpy and zero point energies were acquired from the unscaled vibrations. Single-point energies were calculated at the B3LYP/6-311++G** level and added to their respective enthalpy corrections to give ΔH_{298}. These enthalpy corrections take into account the integrated heat capacity, zero-point energies, and the pressure-volume work term to give the enthalpies of the structures at room temperature.

Chapter 4 – Thermochemistry of Non-Protein Amino Acids

4.1 Introduction

The proton affinities of homocysteine and homoserine were determined by Cook's extended kinetic method in an electrospray ionization quadrupole ion trap instrument. The kinetic method is based on the competitive decomposition of ion-bound dimers by either metastable decomposition or collision-induced dissociation.

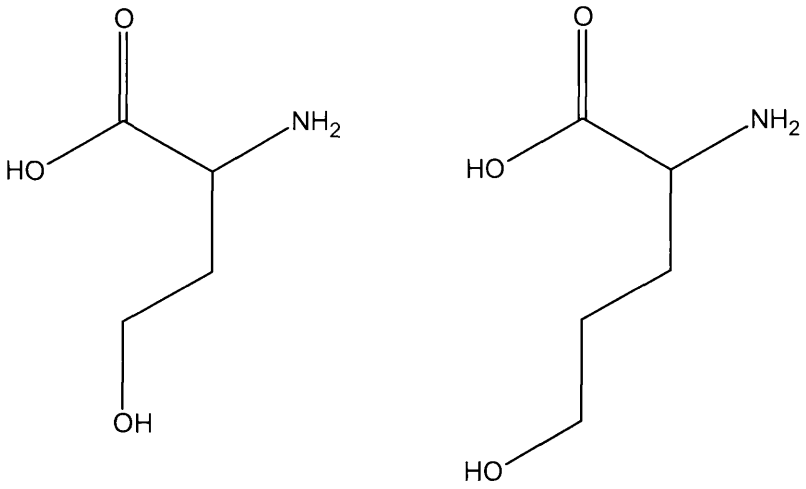
Computational calculations were also conducted with the purpose of obtaining theoretical proton affinities and gas-phase acidities of homocysteine (2-amino-4-sulfanylbutanoic acid, **hcys**), homoserine (2-amino-4-hydroxybutanoic acid, **hser**), homo homocysteine (2-amino-5-sulfanylpentanoic acid, **hhcys**) and homo homoserine (2-amino-5-hydroxypentanoic acid, **hhser**), as can be seen below in Figures 4.1 and 4.2.



Homocysteine

Homo homocysteine

Figure 4.1 – Neutral Structures of Homocysteine and Homo homocysteine



Homoserine

Homo homoserine

Figure 4.2 – Neutral Structures of Homoserine and Homo homoserine

Predictions for the proton affinities and gas-phase acidities were found at 298K from the isodesmic reactions with the reference glycine (PA = 886.5 kJ/mol)[34] and acetic acid (ΔH_{acid} = 1456 kJ/mol)[34], respectively. As shown below in equations(16) and (17).

$$A + {}^{+}H_3NCH_2COOH \longrightarrow AH^{+} + H_2NCH_2COOH \quad (16)$$

$$HA + OAc^{-} \longrightarrow A^{-} + HOAc \quad (17)$$

Calculations were performed for every deprotonation or protonation site present for each NPAA. The lowest-energy optimized geometries were examined to predict deprotonation and protonation sites and detect any structural changes upon proton transfer.

4.2 Thermochemical Properties of Cysteine Homologs

4.2.1 – Methods of Determining Proton Affinities for Cysteine Homologs

The proton affinity of homocysteine was analyzed using Cooks' extended kinetic method. As this is the first study pertaining to the proton affinity of homo amino acids, a trial and error method was utilized in order to narrow down the possible reference bases. Butylamine (PA = 921.5 kJ/mol),[34] when paired with homocysteine, produced an acceptable intensity ratio which allowed it to serve as the starting point for choosing additional reference bases. After selecting against mass discrimination and stability in the ion trap, four more reference bases were chosen. The reference bases and their proton affinities[34] listed in Table 4.1 were used in the determination of the proton affinities of homocysteine. Figure 4.3 illustrates the protonated references bases chosen for use in the determination of the experimental proton affinity for homocysteine.

Table 4.1 – Reference Bases for Homocysteine

Reference Base	**Molecular Weight** (g/mol)	**PA** (kJ/mol)
Benzylamine	107.15	913.3
Propylamine	59.11	917.8
Butylamine	73.14	921.5
Hexylamine	101.19	927.5
Cyclohexylamine	99.17	934.4

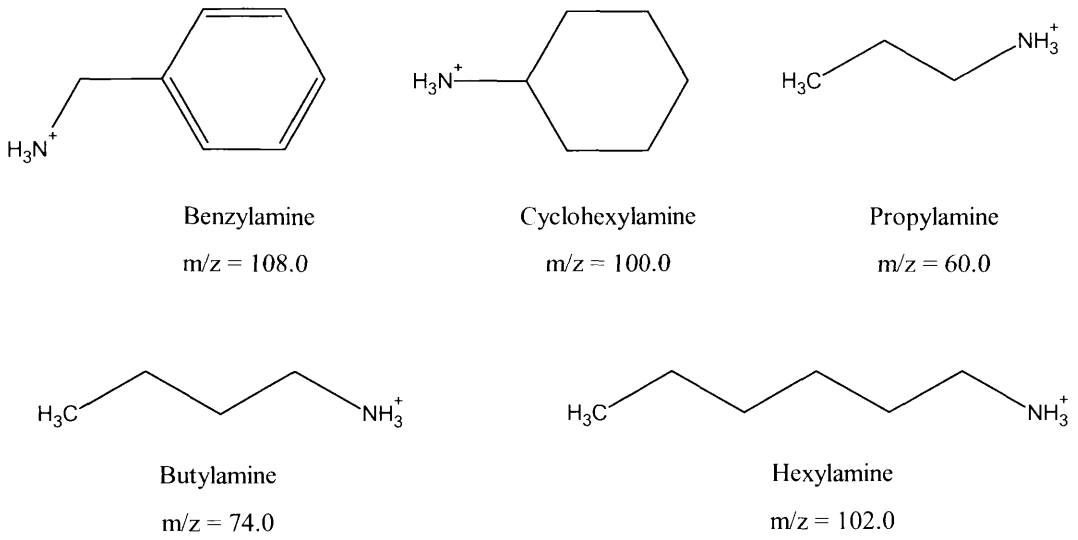


Figure 4.3 – Structures of Protonated Reference Bases for Homocysteine

The homocysteine-reference base fragmentation spectra were collected as described in the procedure in Section 3.1. Using Cooks' extended method, the proton affinity of homocysteine can be obtained as described in Section 2.1. Cooks' extended kinetic method equation will yield a straight line with a slope of $\frac{1}{\mathrm{RT_{eff}}}$ and an intercept given by $-[((\Delta \mathrm{H\_A} - \Delta \mathrm{H\_avg}))/(\mathrm{RT\_eff}) + \Delta\Delta S/R]$ when the intensity ratio, $\ln\left(\frac{\mathrm{B}_i\mathrm{H}^+}{\mathrm{AH}^+}\right)$, is plotted against $\Delta \mathrm{H}_{\mathrm{B}_i} - \Delta \mathrm{H}_{\mathrm{avg}}$. An effective temperature plot was constructed in order to select the range of activation amplitudes chosen for the construction of kinetic method plot 2. Values were chosen in which the effective temperature increased with activation amplitude. A plot of the negative of each of the intercepts versus the slopes from the lines in plot 1 gives kinetic method plot 2. From this plot the proton affinity and a prediction for the entropy of protonation can be obtained (slope = $\Delta \mathrm{H_A} - \Delta \mathrm{H_{avg}}$; y-intercept = $\frac{\Delta\Delta S}{R}$).

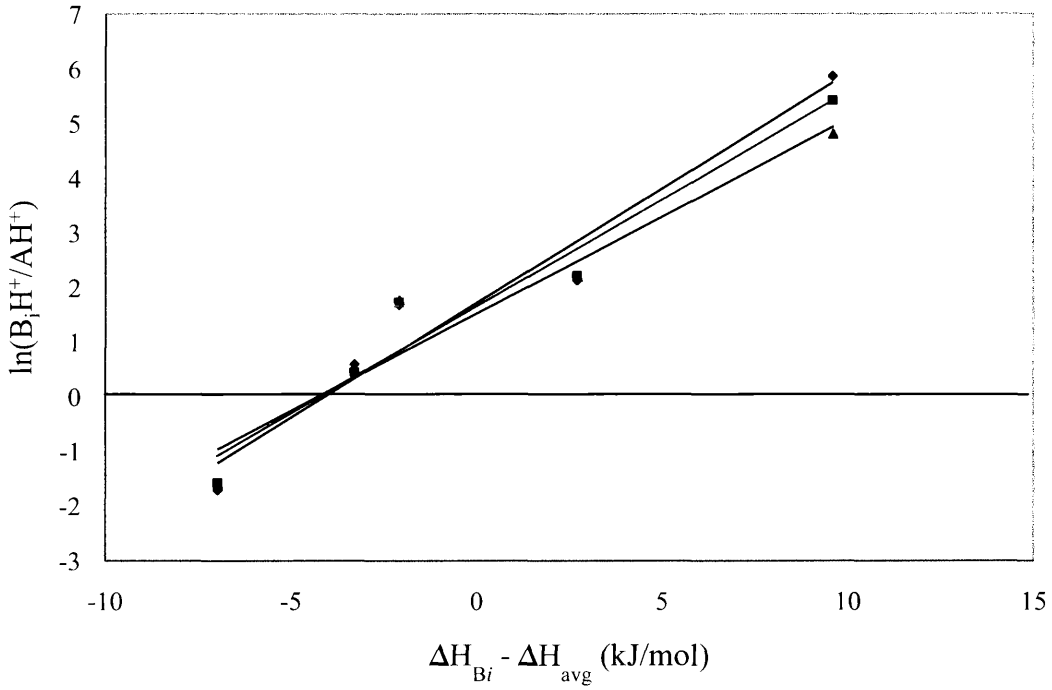


Figure 4.4 – Kinetic Method Plot 1: Homocysteine

Figure 4.4, an illustration of kinetic method plot one, shows a selection of three representative activation amplitudes from the entire scanned amplitude range. Each of the regression lines represents a different activation energy. From the x-axis, the proton affinity for each reference base is represented by a horizontally-equivalent set of three points which correspond to a value ± x from the average basicity of all reference bases. The lines intersect at the isothermal point, where the fragmentation ratio remains constant for all effective temperatures. With an average proton affinity of the reference bases being 924.8 kJ/mol and the isothermal point occurring at -3.9 kJ/mol, an estimate of the extended kinetic method measurement for the proton affinity of homocysteine is 920.9 kJ/mol.

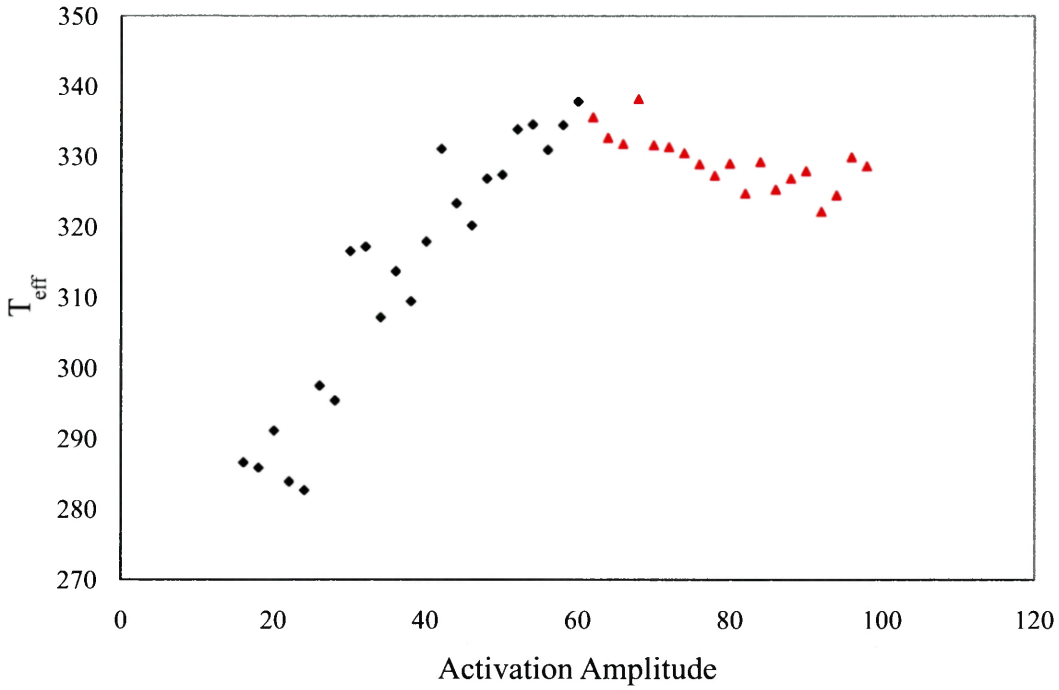


Figure 4.5 – Effective Temperature Plot: Homocysteine

From the effective temperature plot (Figure 4.5), the range of activation amplitudes from 16% to 60% was chosen to construct the second kinetic method plot. Beyond 60%, the effective temperatures no longer increase at a constant rate with the increase in activation amplitude. This effect is attributable to the equilibration of the rate of heating to the rate of cooling as a result of the helium gas.

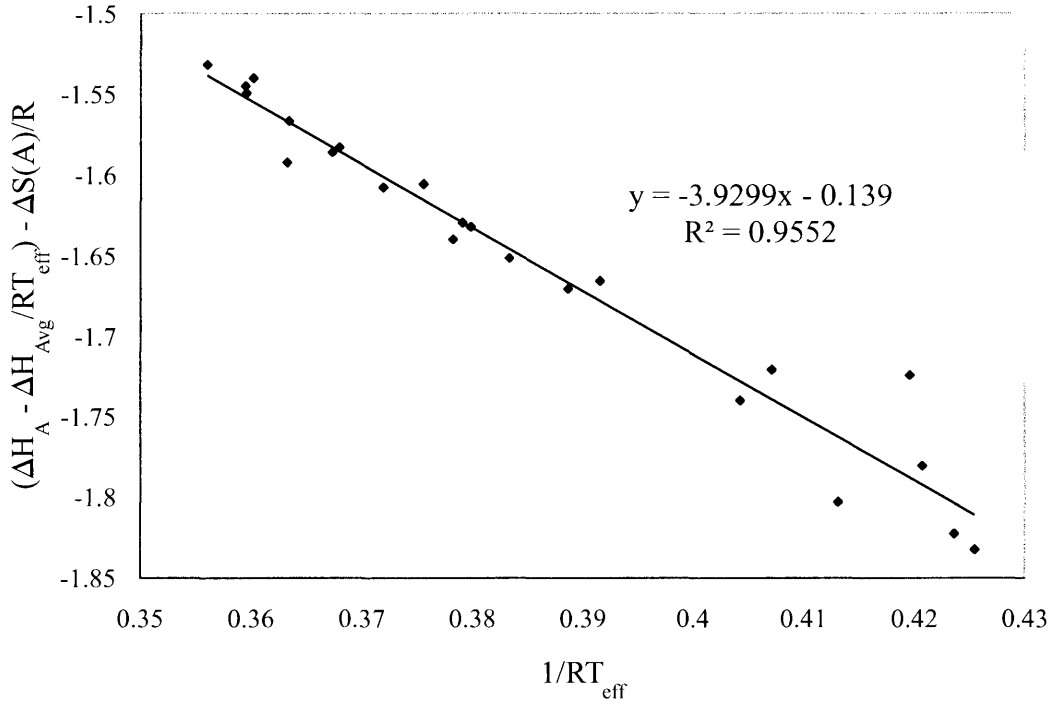


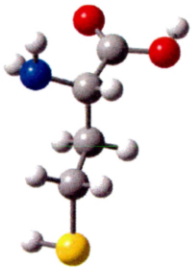
Figure 4.6 – Kinetic Method Plot 2: Homocysteine

As shown by the R^2 value in Figure 4.6, the slope and intercept of the data is accurately reflected by the regression equation. From Cooks' extended kinetic Method with ODR, the proton affinity was measured to be 919.8 ± 12.8 kJ/mol.

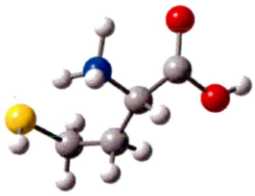
In order to better understand the kinetic method results, theoretical calculations were performed using hybrid density functional theory B3LYP method as described in Section 3.2. Predictions for the proton affinities of homocysteine (PA = 931.7 kJ/mol) and homo homocysteine (PA = 945.2 kJ/mol) were calculated using the isodesmic reaction with glycine (equation 16). The uncertainty in these values is hard to quantify but it is certainly no better than the uncertainties in the isodesmic reference compounds which are on the order of ±8 kJ/mol. We assign error bars of ±10 kJ/mol for the absolute proton affinities and gas-phase acidities. Relative values,

for the various sites in the same molecule, are better than this, probably ±5 kJ/mol.

Figures 4.7 and 4.8 show the lowest energy conformer for neutral and protonated homocysteine and homo homocysteine.

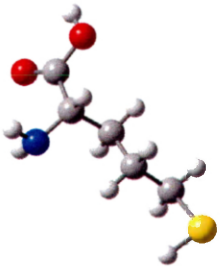


Neutral Homocysteine

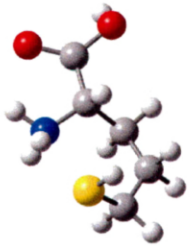


Homocysteine
Protonated on Nitrogen

Figure 4.7 – Geometrically Optimized Neutral and Protonated Homocysteine Structures



Neutral Homo
homocysteine



Homo homocysteine
Protonated on Nitrogen

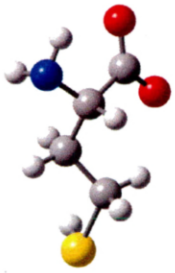
Figure 4.8 – Geometrically Optimized Neutral and Protonated Homo homocysteine Structures

A more extended structure can be seen for the lowest energy conformers of both homocysteine and homo homocysteine while the protonated forms are more compact due to internal hydrogen bonding.

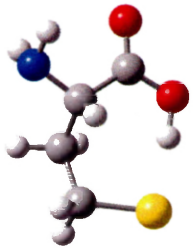
4.2.2 – Methods of Determining Gas-Phase Acidities for Cysteine Homologs

Attempts to calculate the gas-phase acidity utilizing the kinetic method were made. However, the proton bound dimer was not visible while operating the ion trap in negative ion mode. This was due to the inability to condition the instrument for operation in the negative ion mode.

Similar calculations were employed to give predictions for the gas-phase acidities for homocysteine and homo homocysteine utilizing the isodesmic reaction with acetic acid (equation 17). The gas-phase acidities of deprotonated homocysteine and homo homocysteine (Figures 4.9 and 4.10) were calculated to be 1398.5 ± 10.0 kJ/mol and 1408.5 ± 10.0 kJ/mol, respectively.

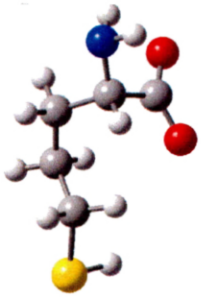


Homocysteine Carboxylate Anion

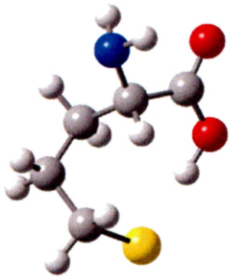


Homocysteine Thiolate Anion

Figure 4.9 – Geometrically Optimized Deprotonated Homocysteine Structures



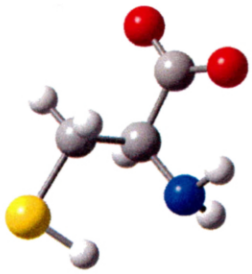
Homo homocysteine
Carboxylate Anion



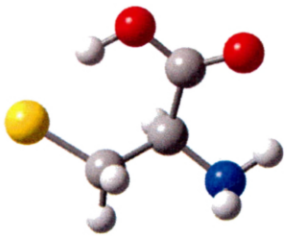
Homo homocysteine
Thiolate Anion

Figure 4.10 – Geometrically Optimized Deprotonated Homo homocysteine Structures

In a recent study by Poutsma and coworkers, the gas-phase acidity of cysteine was determined. It was reported that the thiolate anion was predicted to be 8.0 kJ/mol more stable than the carboxylate anion (Figure 4.11). This result was unexpected due to the fact that carboxylic acids are generally more acidic than thiols. It has been suggested that in the gas phase, intramolecular hydrogen bonding in cysteine overcomes the acidity of carboxylic acids.[35] A comparison of the results from the current study and the study by Poutsma and coworkers[26] can be seen in Table 4.2.



Cysteine Carboxylate Anion*



Cysteine Thiolate Anion*

Figure 4.11 – Geometrically Optimized Deprotonated Cysteine Structures
*Structures taken from Reference [26]

Table 4.2 – Absolute Gas-Phase Acidities of Cysteine Analogs

Molecule	Calculated Gas-Phase Acidity (kJ/mol)
Cysteine (carboxylate)	1394.0 ± 10.0*
Cysteine (thiolate)	1386.0 ± 10.0*
Homocysteine (carboxylate)	1398.8 ± 10.0
Homocysteine (thiolate)	1397.2 ± 10.0
Homo homocysteine (carboxylate)	1412.0 ± 10.0
Homo homocysteine (thiolate)	1408.5 ± 10.0

*Values taken from Reference [26]

It was found that the homocysteine and homo homocysteine results parallel those reported by Poutsma and coworkers in the cysteine study in that the thiol group is slightly more acidic than the acid group. Although, given the uncertainties of these values, it is more proper to say that the two sites have equal acidities in both homocysteine and homo homocysteine.

4.3 Thermochemical Properties of Serine Homologs

4.3.1 – Methods of Determining Proton Affinities for Serine Homologs

Reference bases for homoserine were chosen from a modified homocysteine reference base set. The bases chosen for use in the kinetic method study and their proton affinities[34] are displayed in Table 4.3. Figure 4.11 illustrates the protonated reference bases chosen for use in the experimental determination of the proton affinity of homoserine.

Table 4.3 – Reference Bases for Homoserine

Reference Base	**Molecular Weight** (g/mol)	**Proton Affinity** (kJ/mol)
Butylamine	73.14	921.5
Hexylamine	101.19	927.5
Pyridine	79.10	930.0
3-Picoline	93.13	943.4

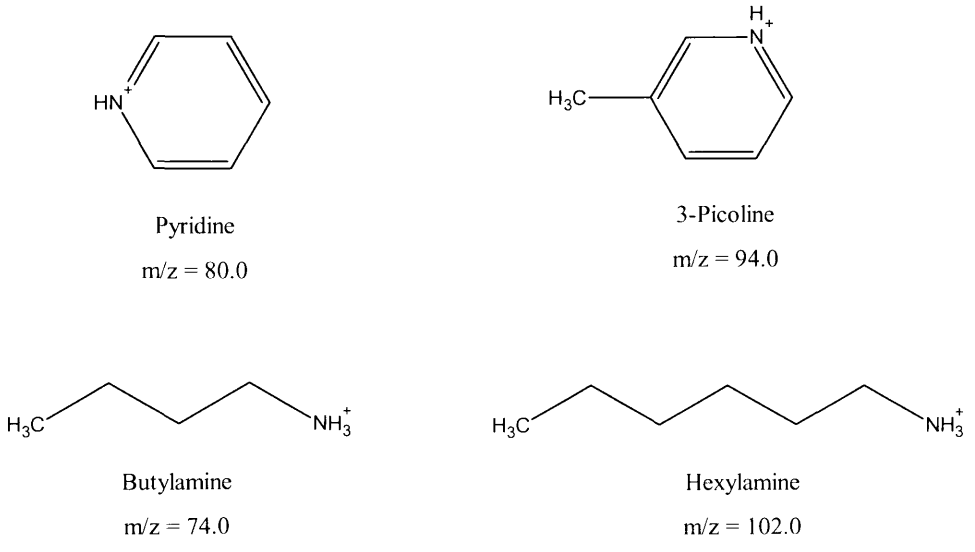


Figure 4.12 – Structures of Protonated Reference Bases for Homoserine

Homoserine and reference base solutions were prepared in a 49.5:49.5 methanol:water mixture containing 1% acetic acid. Again, the heterodimer peak was isolated and the activation amplitude scanned from 2% to 100%. The fragmentation ratios at each of the amplitudes were recorded and kinetic method plots one and two and an effective temperature plot were constructed for the representative activation amplitudes.

Kinetic method plot one for homoserine can be seen in Figure 4.12. The average proton affinity of the reference bases was found to be 930.6 kJ/mol with the isothermal point occurring at 7.7 kJ/mol. From these values, an estimate of the extended kinetic method measurement for the proton affinity of homoserine was calculated to be 938.3 kJ/mol.

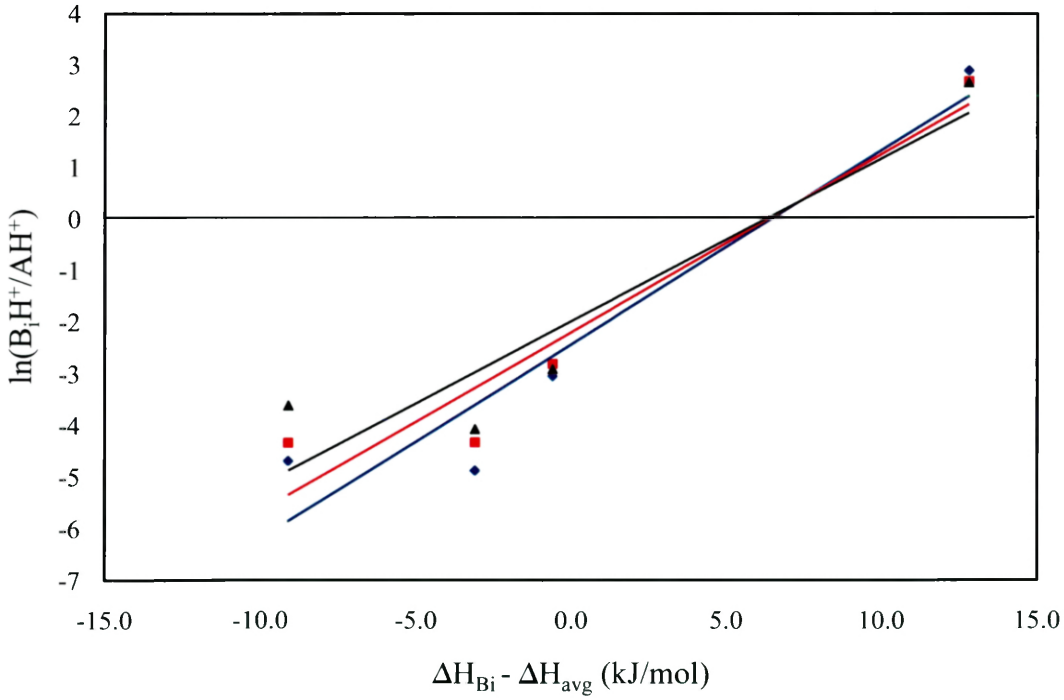


Figure 4.13 – Kinetic Method Plot 1: Homoserine

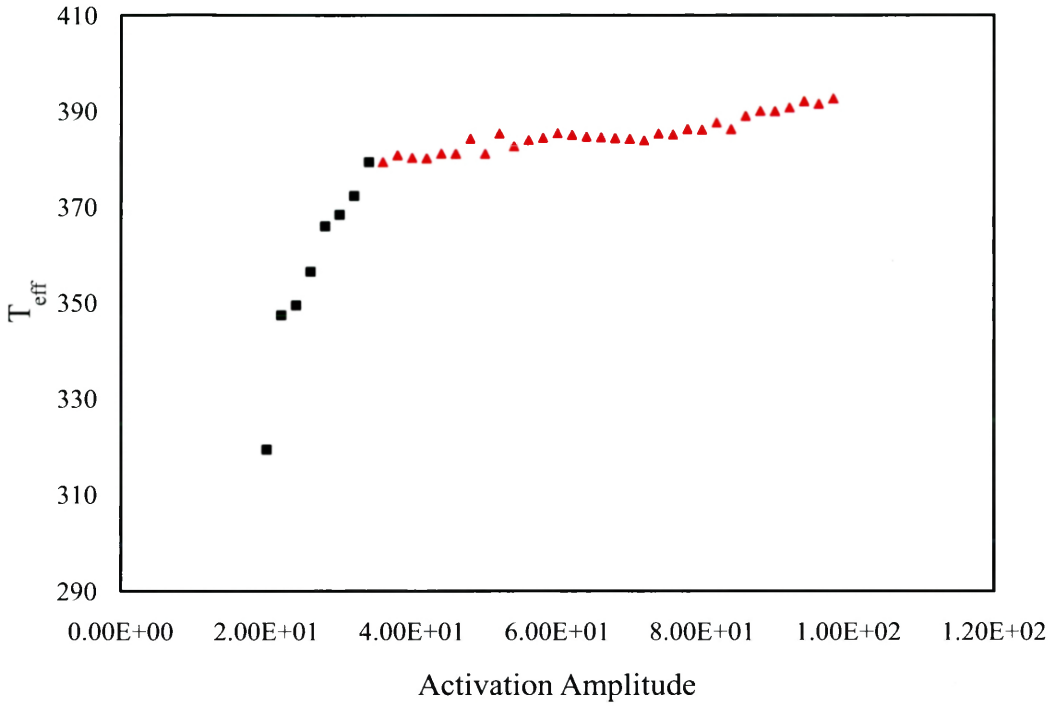


Figure 4.14 – Effective Temperature Plot: Homoserine

Based on the effective temperature plot, Figure 4.13, a range of activation amplitudes from 20 % to 34% was chosen to generate the second kinetic method plot.

The slope and intercept of the data is accurately reflected by the regression equation as can be seen by the R^2 value in Figure 4.14. From Cooks' extended kinetic method with ODR, the proton affinity was measured to be 942.5 ± 18.0 kJ/mol.

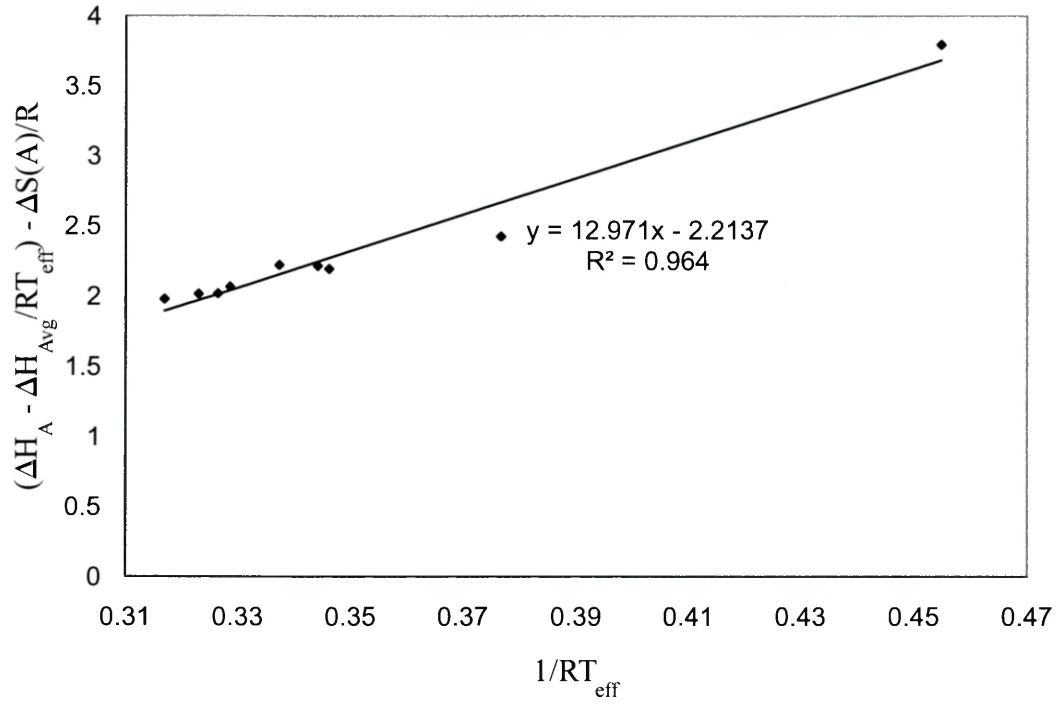
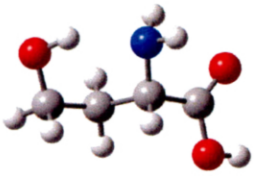
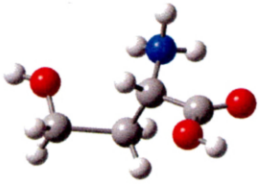


Figure 4.15 – Kinetic Method Plot 2: Homoserine

Hybrid density functional theory calculations were again performed to give predictions for the proton affinities of homoserine and homo homoserine. Using the isodesmic reaction with glycine (equation 16), the proton affinities were determined to be 942.1 kJ/mol and 975.0 kJ/mol, respectively. Figures 4.15 and 4.16 show the lowest energy conformer for neutral and protonated homoserine and homo homoserine.

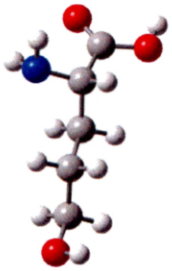


Neutral Homoserine

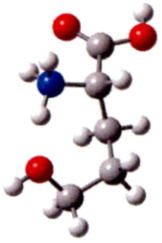


Homoserine Protonated on Nitrogen

Figure 4.16 – Geometrically Optimized Neutral, and Protonated Homoserine Structures



Neutral Homo homoserine



Homo homoserine Protonated on Nitrogen

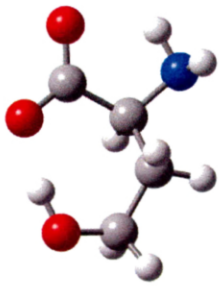
Figure 4.17 – Geometrically Optimized Neutral and Protonated Homo homoserine Structures

As was seen with the cysteine analogs, the neutral structures for homoserine and homo homoserine exhibit an extended arrangement while the protonated structures show evidence of contraction attributable to additional hydrogen bonding.

4.3.2 – Methods of Determining Gas-Phase Acidities for Seine Homologs

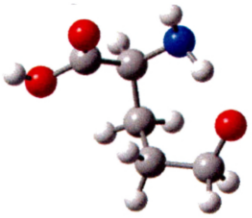
As with homocysteine, an experimental kinetic method determination of the gas-phase acidity of homoserine was not possible due to the inability to form the proton-bound dimers. The isodesmic reaction with acetic acid (equation 17), was again utilized in order to calculate the gas-phase acidities of homoserine and homo homoserine. These values were found to be 1397.2 ± 10.0 kJ/mol and 1408.5 ± 10 kJ/mol, respectively.

Figures 4.18 and 4.19 illustrate the geometrically optimized structures used in the determination of the gas-phase acidities of homoserine and homo homoserine. We were unable to locate an optimized structure for homoserine deprotonated on the alcohol. All conformers that started out deprotonated at the alcohol isomerized to conformers deprotonated at the carboxylic acid by proton transfer during geometry optimization. Consequently, only the acid deprotonated structure is shown in Figure 4.18.

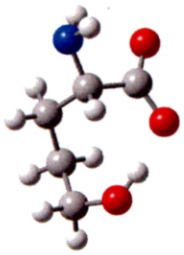


Homoserine Deprotonated
on Acid

Figure 4.18 – Geometrically Optimized Deprotonated Homoserine Structures



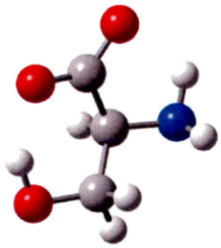
Homo homoserine
Deprotonated on
Alcohol



Homo homoserine
Deprotonated on
Acid

Figure 4.19 – Geometrically Optimized Deprotonated Homo homoserine Structures

A recent study by Poutsma and coworkers has investigated the intrinsic properties of serine.[26] It has been reported that the serine structure deprotonated on the acid is more stable than the structure deprotonated on the alcohol (Figure 4.20). A comparison of the results from the current study and the study by Poutsma and coworkers[26] can be seen in Table 4.4.



Serine Deprotonated
on Acid*

Figure 4.20 – Geometrically Optimized Deprotonated Serine Structures
*Structure taken from Reference [26]

Table 4.4 – Absolute Gas-Phase Acidities of Serine Analogs

Molecule	Calculated Gas-Phase Acidity (kJ/mol)
Serine Deprotonated on Acid	1382.3 ± 10.0*
Homoserine Deprotonated on Acid	1393.6 ± 10.0
Homo homoserine Deprotonated on Alcohol	1524.1 ± 10.0
Homo homoserine Deprotonated on Acid	1383.4 ± 10.0

*Values taken from Reference [26]

In serine and its homologs, it was observed that the carboxylate anion was more stable than that of the hydroxylate anion. As was seen with homoserine, an optimized structure for serine deprotonated on the alcohol was not located. As can be seen from the table, the effect of lengthening the side chain of serine on its gas-phase acidity is small in that serine, homoserine, and homo homoserine all have the same acidity within error.

4.4 Discussion

4.4.1 – Cysteine Homologs

The proton affinities and gas-phase acidities of cysteine analogs were measured in order to determine whether trends corresponding to the side chain length were present.

Proton-bound dimers of homocysteine and one of a series of reference bases were generated from electrospray ionization. The following reference bases were used: benzylamine, propylamine, butylamine, hexylamine, and cyclohexylamine. The recommended proton affinity for this compound was found to be 919.8 ± 12.8

kJ/mol. The calculated proton affinity for homocysteine, 931.7 ± 10 kJ/mol, falls within the experimental error of the measured proton affinity.

Table 4.5 displays the results for the computational proton affinity and gas-phase acidity determinations as compared to the experimental data using Cooks' extended kinetic method with ODR. An experimental proton affinity is not available for homo homocysteine due to the inability to obtain the commercially produced molecule.

Table 4.5 – Comparison of Computational Studies and Experimental Studies for Cysteine Analogs

	Experimental Proton Affinity (kJ/mol)	**Calculated Proton Affinity** (kJ/mol)	**Calculated Gas-Phase Acidities** (kJ/mol)
Cysteine	903.3*	903.2**	1393.0 ± 13.0***
Homocysteine	919.8 ± 12.8	931.7 ± 10.0	1397.2 ± 10.0
Homo homocysteine	–	945.2 ± 10.0	1408.5 ± 10.0

*Values taken from Reference [36]
**Values taken from Reference [34]
***Values taken from Reference [37]

Hybrid density functional theory calculations have predicted the proton affinity and gas-phase acidity of cysteine analogs. Theoretical thermochemical values for the lowest energy conformers of neutral and protonated species investigated in this work are listed in Table 4.6. For each analog, 40 different conformations were investigated.

Table 4.6 – Total Electronic Energies, Thermal Corrections and 298K Enthalpies (Hartrees) for Selected Cysteine Analogs

Molecule	total electronic energy[a]	thermal correction[b]	ΔH_{298}
Homocysteine	-761.3928221	0.14728	-761.2455421
Homocysteine Deprotonated on Hydroxyl	-760.8538618	0.134924	-760.7189378
Homocysteine Deprotonated on Sulfur	-760.8535279	0.135507	-760.7170209
Homocysteine Protonated on Nitrogen	-761.7578423	0.161371	-761.5964713
Homo homocysteine	-800.7174070	0.177169	-800.5402380
Homo homoserine Deprotonated on Hydroxyl	-800.1749889	0.165148	-800.0098409
Homo homocysteine Deprotonated on Sulfur	-800.1704378	0.164494	-800.0059438
Homo homocysteine Protonated on Nitrogen	-801.0862706	0.189963	-800.8963076

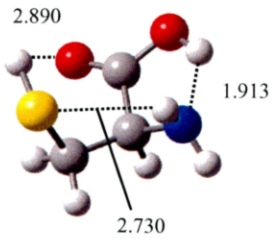
[a] B3LYP/6-311++G**//B3LYP/6-31+G* [b] ZPE + $\int c_p dT$, from unscaled harmonic frequencies obtained at B3LYP/6-31+G*

The importance of the amino acid side chain is most commonly attributed to its thermodynamic properties. In addition, the length of the side chain affects the number of possible fragmentation pathways. A general model for how peptides fragment in the gas phase, termed the mobile proton model, has been developed. The mobile proton heavily influences peptide fragmentation pathways and is strongly affected by the amino acids that make up the peptide. When a side chain includes a group that is particularly basic, the proton is usually sequestered at that site allowing for site directed fragmentation. Shortening the side chain length limits the possible fragmentation pathways as well as the possible hydrogen bonding sites. A more thorough discussion of the mobile proton model can be found in Chapter 5.

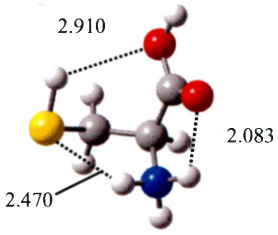
Hunter and Lias[34] have reported the absolute proton affinity for cysteine to be 903.2 kJ/mol. The hybrid DFT calculations performed in this study in conjunction

with the study of Hunter and Lias illustrates the correlation of side chain length with thermochemical properties. The addition of a methylene to cysteine produces an increase of 29.1 kJ/mol in the proton affinity. However, the addition of a second methylene only produces an increase of 12.9 kJ/mol. Although the proton affinity shows an overall increase as the length of the side chain increases, the addition of the second methylene gives an increase only half as large as seen with the addition of the first methylene. Additionally, the gas-phase acidities exhibit a decrease in acidity by 5.5 and 11.3 kJ/mol, respectively.

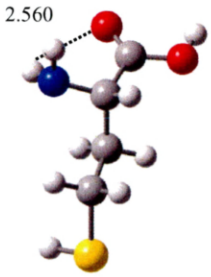
Hydrogen bonding plays an important role in determining physical and chemical properties of peptides. The contributions of hydrogen bonding to stabilize conformations are reflected in the energy of the molecule and as well as the predicted proton affinity. Cysteine analogs and their conjugate acids were fully optimized at the B3LYP/6-311++G** level. Figures 4.21 and 4.22 illustrate the intramolecular hydrogen bonding schemes found in these molecules.



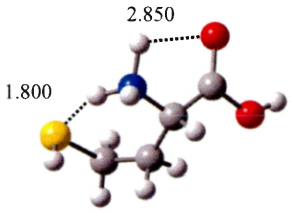
Neutral Cysteine*



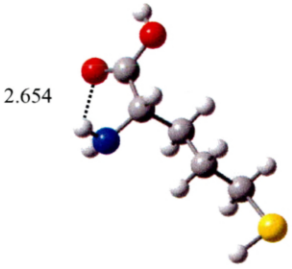
Cysteine Protonated on Nitrogen**



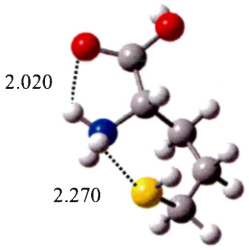
Neutral Homocysteine



Homocysteine Protonated on Nitrogen



Neutral Homo homocysteine

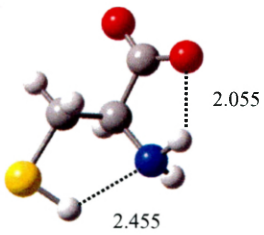


Homo homocysteine Protonated on Nitrogen

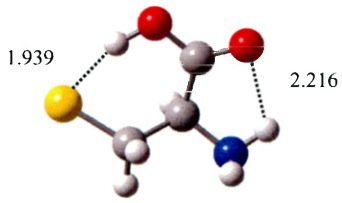
Figure 4.21 – B3LYP/6-311++G** Optimized Structures and Hydrogen Bond Length for Cysteine Analogs and the Conjugate Acids

*Structure obtained from reference [38]

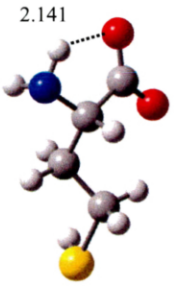
** Structure obtained from reference [36]



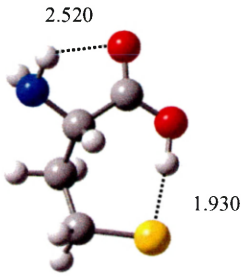
Cysteine Carboxylate Anion*



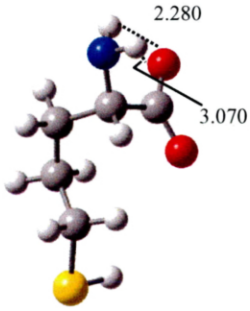
Cysteine Thiolate Anion*



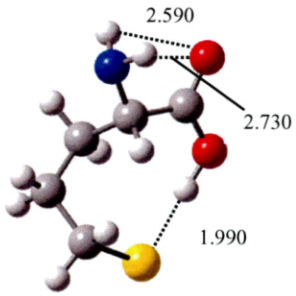
Homocysteine Carboxylate Anion



Homocysteine Thiolate Anion



Homo homocysteine Carboxylate Anion



Homo homocysteine Thiolate Anion

Figure 4.22 – B3LYP/6-311++G** Optimized Structures and Hydrogen Bond Length for the Conjugate Bases of the Cysteine Analogs
*Structures taken from Reference [26]

The lowest energy structure for neutral cysteine involves three hydrogen bonds; one between the carboxyl group and the amine (CO_2H•••N), the second between the thiol and the carbonyl of the carboxyl group (SH•••O=C), and the third between the amine and the thiol (NH•••S). Both neutral homocysteine and homo homocysteine are extended and possess one intramolecular hydrogen bond between NH•••O=C.

Protonated cysteine contains three hydrogen bonds; one between NH•••S, a second between SH•••CO_2H, and the third between NH•••O=C. The lowest energy conformers for both protonated homocysteine and protonated homo homocysteine experience hydrogen bonding between NH•••O=C and NH•••S.

The carboxylate anion of cysteine contains two hydrogen bonds; one between NH•••CO_2^- and the second between SH•••N, while the thiolate anion contains hydrogen bonds between OH•••S and NH•••O=C. Deprotonated homocysteine and homo homocysteine possess the same hydrogen bonding schemes in which the carboxylate anion contains two hydrogen bonds between NH•••CO_2^- and the thiolate anion contains two hydrogen bonds; one between the NH•••CO_2^- and the second between OH•••S.

A comparison between the hydrogen bonding schemes of the neutral homologs and the thiolate anion homologs illustrate that in all cases except for cysteine, the deprotonated homologs take on a more constricted form. As mentioned earlier, Poutsma and coworkers have suggested that when the thiolate anion is created, a set of favorable hydrogen bonds is formed. Whereas when the carboxylate anion is formed, the hydrogen bonds are not as strong thus producing a less stable

structure.[35] As with cysteine, this is observed in both homocysteine and homo homocysteine. Regardless of the extended structures, the deprotonation at the sulfur is energetically more favorable.

Recent studies by Poutsma and coworkers on the intrinsic properties of lysine homologs showed similar trends to those seen in the cysteine homologs. 2,3-diaminopropanoic acid (DAPA), 2,4-diaminobutanoic acid (DABA), and ornithine are the lysine homolog equivalent to cysteine, homocysteine, and homo homocysteine. Structures of the lysine homologs can be seen in Figure 4.23.

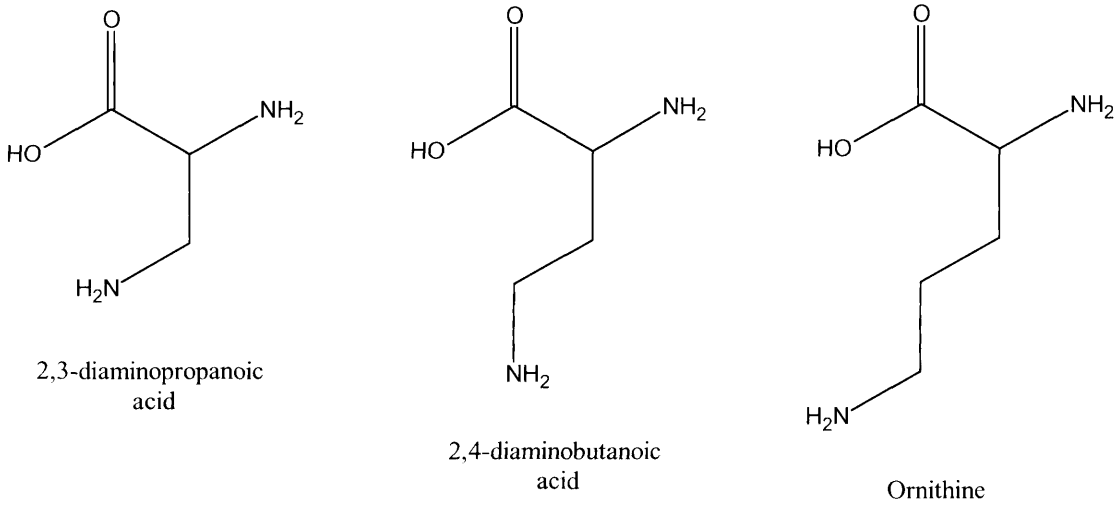


Figure 4.23 – Neutral Structures of Lysine Analogs[15]

Proton affinities of 950.2, 975.8, and 1001.1 kJ/mol were reported for the lysine homologs, respectively. An increase in the proton affinity of 25.6 kJ/mol was seen for the addition of the first methylene group. The addition of the second methylene group also produced an increase in the proton affinity of 25.3 kJ/mol.

Furthermore, a decrease of 5.5 and 11.3 kJ/mol were observed in the gas-phase basicities for the addition of the two methylene groups.[15]

The less electronegative sulfur of cysteine has a higher charge capacity than that of the oxygen of serine, as can be seen in Section 4.4.2. The proton affinities of these homologs are in agreement with this statement. This trend is continued throughout all the homologs; however, a significant increase is seen between homo homocysteine and homo homoserine. Also, as the chain length increases, the proton affinity is shown to increase. The molecule becomes more unstable with the increase in chain length. As in the lysine homologs, it appears that the inductive effects are overwhelmed by the stabilization of the cation by the intramolecular hydrogen bonding.

4.4.2 – Serine Homologs

Similar procedures were carried out to determine the proton affinity of homoserine. The following reference bases were used: butylamine, hexylamine, pyridine, and 3-picoline. The recommended proton affinity for this compound was found to be 942.5 ± 18.0 kJ/mol. This calculated proton affinity for homoserine, 942.1 ±10 kJ/mol, is in excellent agreement with the measured proton affinity.

Table 4.7 displays the results for the computational proton affinity determinations as compared to the experimental data using Cooks' extended kinetic method with ODR. Again, an experimental proton affinity is not available for homo homoserine due to the inability to obtain the commercially produced molecule.

Table 4.7 – Comparison of Computational Studies and Experimental Studies for Serine Analogs

	Experimental Proton Affinity (kJ/mol)	Calculated Proton Affinity (kJ/mol)	Calculated Gas-Phase Acidities (kJ/mol)
Serine	912.5*	914.6**	1392.0 ± 13.0***
Homoserine	942.5 ± 18.0	942.1 ± 10.0	1393.6 ± 10.0
Homo homoserine	–	975.0 ± 10.0	1383.4 ± 10.0

*Values taken from Reference [36]
**Values taken from Reference [34]
***Values taken from Reference [37]

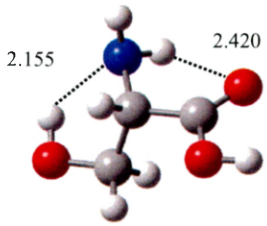
Density functional theory calculations were also carried out for the serine analogs. As with the calculations of the cysteine analogs, B3LYP/6-311++G** calculations predicted the lowest energy structures for the serine analogs. Theoretical thermochemical values for the lowest energy conformers of neutral and protonated species investigated in this work are listed in Table 4.8.

Table 4.8 – Total Electronic Energies, Thermal Corrections and 298K Enthalpies (Hartrees) for Selected Serine Analogs

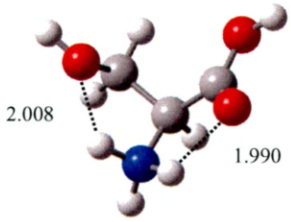
Molecule	total electronic energy[a]	thermal correction[b]	ΔH_{298}
Homoserine	-438.4233874	0.152319	-438.2710684
Homoserine Deprotonated on Acid	-437.8846934	0.138352	-437.7463414
Homoserine Protonated on Nitrogen	-438.7932110	0.166637	-438.6265740
Homo homoserine	-477.7416451	0.181699	-477.55994961
Homo homoserine Deprotonated on Acid	-477.2074731	0.168367	-477.03910610
Homo homoserine Deprotonated on Alcohol	-477.2065705	0.168645	-477.03792550
Homo homoserine Protonated on Nitrogen	-478.1238604	0.196474	-477.9273864

[a] B3LYP/6-311++G**//B3LYP/6-31+G* [b] ZPE + $\int c_p dT$, from unscaled harmonic frequencies obtained at B3LYP/6-31+G*

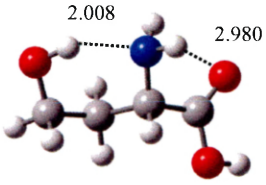
The DFT calculations again show a correlation between the side chain length and the proton affinity. Unlike the cysteine homologs, a uniform increase is seen with the addition of each methylene group. An increase of 27.5kJ/mol and 32.9 kJ/mol was seen for the addition of the first and second methylene groups, respectively. Additionally, an increase in the gas-phase acidity of 1.6 and a decrease of 10.2 kJ/mol was seen with the addition of each methylene group. Although the magnitude changes with the addition of each methylene group, the gas-phase acidities of serine, homoserine, and homo homoserine are all within the assigned error bars of ±10 kJ/mol.



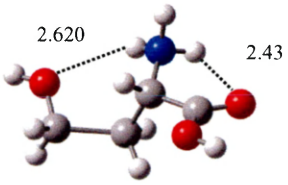
Neutral Serine**



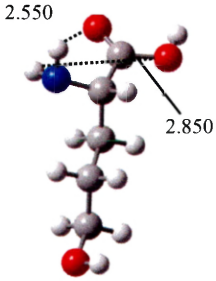
Serine Protonated on Nitrogen*



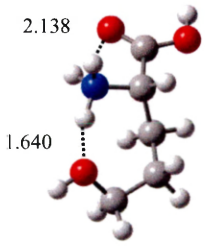
Neutral Homoserine



Homoserine Protonated on Nitrogen



Neutral Homo homoserine

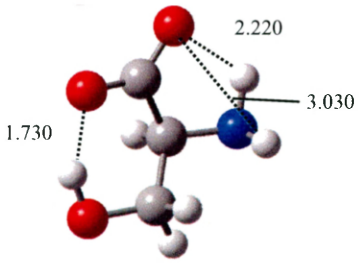


Homo homoserine Protonated on Nitrogen

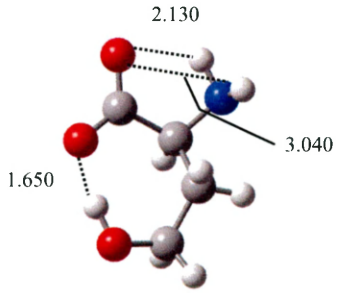
Figure 4.24 – B3LYP/6-311++G** Optimized Structures and Hydrogen Bond Length for Serine Analogs and the Conjugate Acids

*Structure obtained from reference [38]

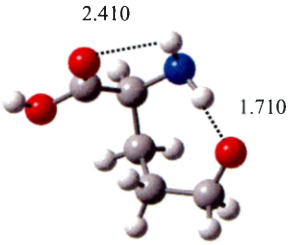
** Structure obtained from reference [36]



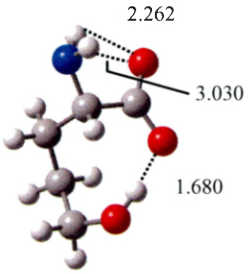
Serine Deprotonated on Acid*



Homoserine Deprotonated on Acid



Homo homoserine Deprotonated on Alcohol



Homo homoserine Deprotonated on Acid

Figure 4.25 – B3LYP/6-311++G** Optimized Structures and Hydrogen Bond Length for the Conjugate Bases of the Serine Analogs
*Structure taken from Reference [26]

Figures 4.24 and 4.25 illustrate the intramolecular hydrogen bonding schemes found in the serine analogs. The lowest energy structure for neutral serine contains two hydrogen bonds; one between the amine and the carbonyl of the carboxyl group (NH•••O=C) and the second between the amine and the hydroxyl group (N•••HO). The neutral homoserine structure includes two hydrogen bonds; the first between NH•••O=C and the second between N•••HO. Homo homocysteine exhibits an

extended structure possessing two intramolecular hydrogen bonds between NH•••O=C and NH•••CO_2H.

Protonated serine contains two hydrogen bonds; one between NH•••O=C and the second between NH•••OH. The lowest energy conformers for both protonated homoserine and protonated homo homoserine experience hydrogen bonding between NH•••CO_2^- and OH•••CO_2^-.

Serine deprotonated on the acid contains three hydrogen bonds; one between OH•••CO_2^- and two between NH•••CO_2^-. Similarly, homoserine and homo homoserine deprotonated on the acid both possess hydrogen bonds between NH•••CO_2^- and OH•••CO_2^-. The structure deprotonated on the alcohol contains bonds between NH•••O=C and NH•••O.

Once more, a comparison can be made with the lysine study by Poutsma and coworkers. A similar increase in proton affinity is seen, however, instead of the decrease in gas-phase acidity seen in both the cysteine and lysine homologs, an increase is observed.

Chapter 5 – Acetylated Proline Analogs

5.1 Protein Identification – Mass Spectrometry

The study of proteomics has recently become a hot topic in the biological sciences and tremendous developments have been made in the understanding of the human proteome. Sophisticated mass spectrometry techniques have replaced older and more time consuming methods such as Edman degradation. In the Edman degradation method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues. Each amino acid is removed in this fashion until the carboxyl-terminus is reached.

Mass spectrometry based sequencing methods have been gaining popularity in recent years as new techniques and increasing computing power have facilitated it. In principle, mass spectrometry can sequence any size protein. However, in practice, it becomes more computationally taxing as the size increases. Peptides are better suited for protein sequencing utilizing mass spectrometry due to not only the size constraints but also solubility issues.

Bottom-up or shotgun proteomics is the most common method of protein sequencing employing mass spectrometry. This method involves the use of an enzyme to digest the protein of interest followed by one or more dimensions of separation by liquid chromatography coupled to mass spectrometry.[39] The sequencing involves the comparison of simulated spectra obtained from computer algorithms with experimental spectra. These computer algorithms use established rules to predict simulated spectra for the protein of interest. Most peptides break in

random positions along the peptide backbone producing a multitude of possible fragments. Unfortunately, certain amino acid residues encourage selective cleavages that can conceal other random cleavages that can aid in peptide identification.[39,41] Examples of these selective cleavages include the "proline effect" in which a preference for cleavage of the amino-terminal to proline residues is seen and the "aspartic acid effect" in which selective cleavage of the amide bond C-terminal is dominant.[42]

Figure 5.1 shows the standard nomenclature for peptide fragmentation developed by Roepstorff and Kohlman. This nomenclature classifies fragment ions as resulting from the breaking of one of the three backbone bonds and whether the charge stays with the amino or carboxyl-terminus. The a_n, b_n, and c_n ion fragments result from the charge staying with the amino-terminal fragment and the x_{m-n}, y_{m-n}, and z_{m-n} ions result from the charge staying with the carboxyl-terminal fragment. The *n* subscript denotes the number or R-groups present on the fragment ion while the *m* subscript represents the total amount of R-groups on the peptide chain.[43,44]

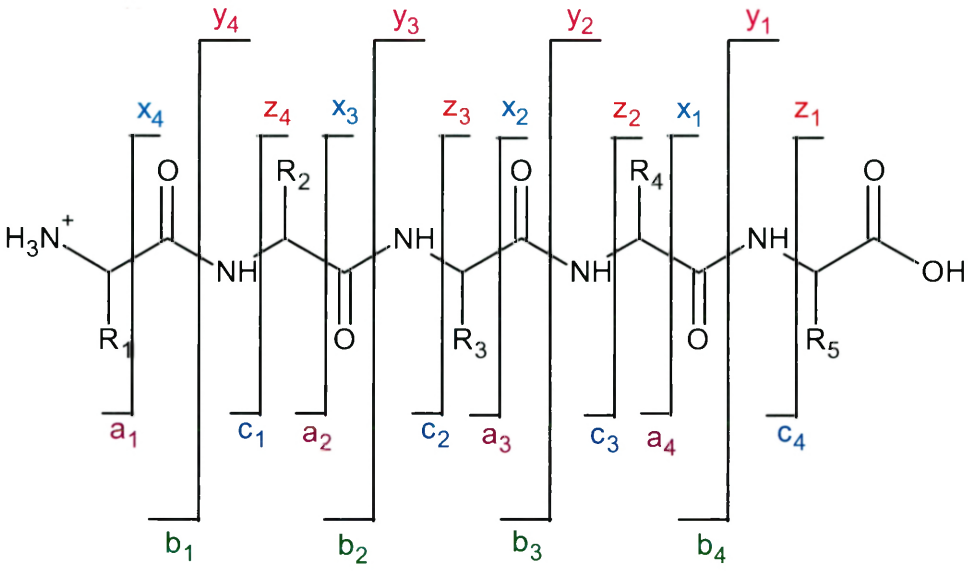


Figure 5.1 – Peptide Fragmentation Nomenclature Diagram

In low energy CID experiments, peptide chains will fragment into b, y, and on occasion, a ions. It is from these three ions that an assortment of fragments and unique fragmentation patterns are formed.

5.2 The "Proline Effect"

Proline-rich proteins, while relatively uncommon, comprise several abundant classes of proteins in both plants and animals. Due to the high efficiency of cleavage at the amide bond on the N-terminal of proline residues and, as a result, the low relative abundance of fragmentation due to cleavage at other amide bonds, proteins containing high amounts of proline are extremely difficult to sequence by mass spectrometry.[45]

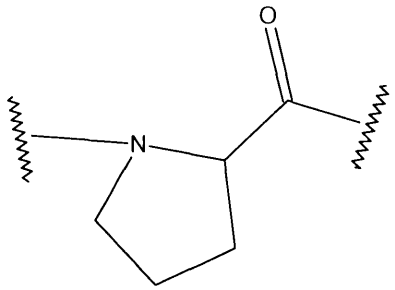


Figure 5.2 – Orientation of Proline Residue in a Peptide

Numerous theoretical and experimental studies have been performed in order to understand the mechanism for the proline effect. Initially, it was proposed that the main cause of the "proline effect" was the enhanced basicity of the proline residue.[46] A few years later, Vaisar and Urban studied the fragmentation patterns of a series of protonated peptides with the general structure of H-Ala-Val-Xxx-Leu-Gly-OH where Xxx is proline, pipecolic acid, or N-methylalanine. In their analysis, they have postulated that the "proline effect" appears to be due to a combination of several factors where the unfavorable structure of the b_3^+ ion plays the major role. The high proton affinity of the proline residue is a secondary factor, which after the discrimination of the formation of the b_3^+ ion appears to determine the abundance of the y_3^+ ion.[47]

Several research groups have contributed to the development of a general model for how peptides fragment in the gas-phase. These studies have helped to define and improve the mobile proton model.[48-50] The mobile proton model is the most comprehensive model currently available in which the dissociation of protonated peptides upon excitation is described. This model has emerged as a result

of a vast number of studies performed by Wysocki and Dongré.[51-55] This mobile proton model is shown below in Figure 5.3

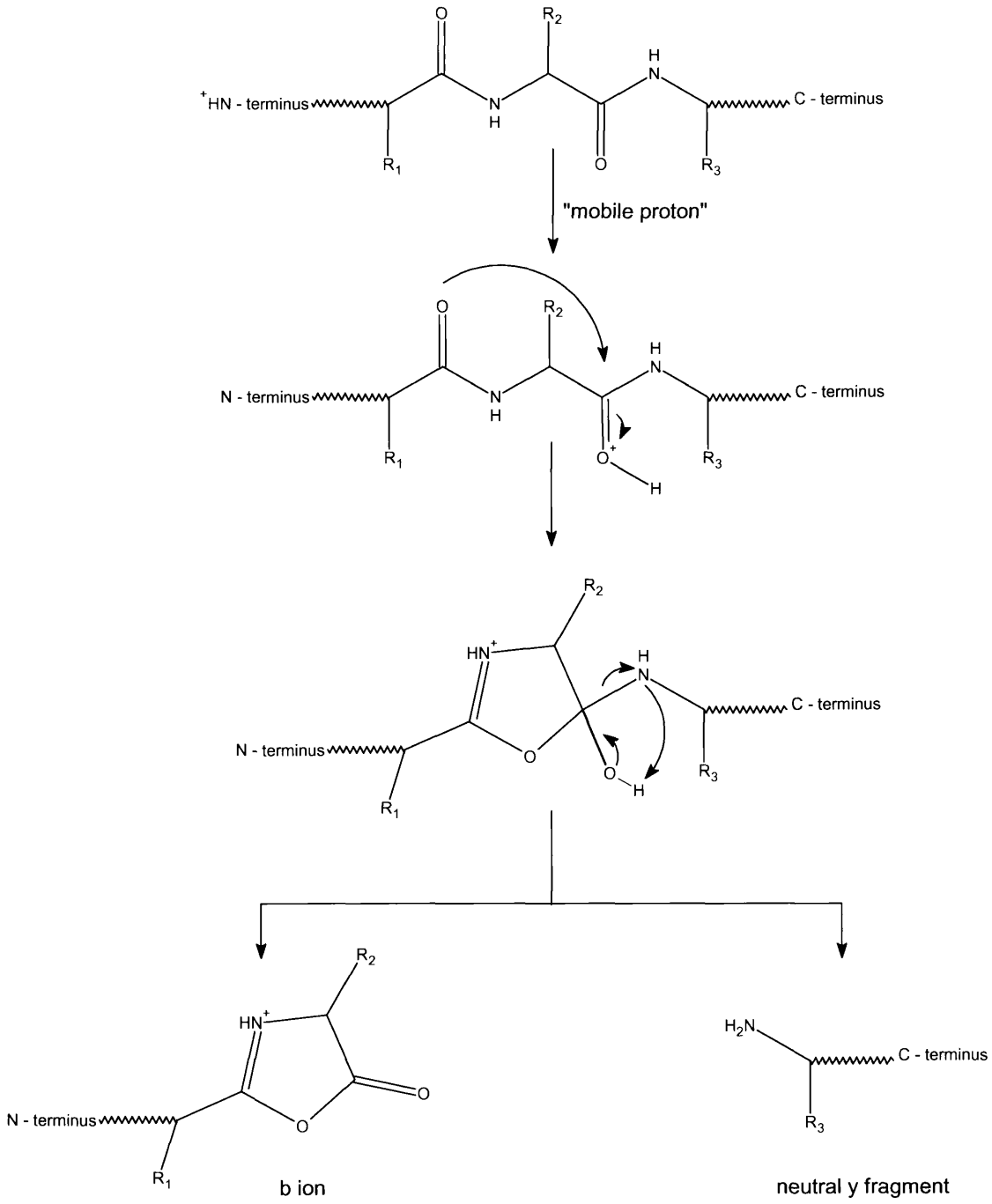


Figure 5.3 – Fragmentation Mechanism Via Protonated Oxazolone Intermediate[56]

Due to the multifunctional nature of peptides, they can be protonated at various sites leading to various isomers. From this, two major classes of peptide ions emerge. In the first class, one or more of the protonation sites is kinetically and/or energetically more favored than the other leading to sequestration of the added proton(s). A large amount of energy is needed to mobilize the extra proton(s) to the energetically less favorable protonation sites. In the second class, many of the protonation sites are accessible in a narrow energy range. Energetically less favorable protonation sites can become more populated or charge-remote fragmentation can occur, as the internal energy of the ions is increased upon excitation. Molecular orbital calculations show substantial weakening of the amide bond upon protonation on the amide nitrogen and protonation on the amide oxygen makes the amide bonds stronger than the neutral species. Alternatively, protonation on the amide oxygen is thermodynamically favored compared to protonation on the amide nitrogen. The mobile proton model resolves this discrepancy by stating that upon excitation, providing the proton(s) are not sequestered by a basic site, they will migrate to various protonation sites prior to fragmentation.[39,54]

Paizs and Suhai[56] have proposed a mechanism (Figure 5.4) for b_3^+ and y_3^+ ion formation from Ala-Gly-Pro-Gly-Ala based on initial transfer of the mobile proton to the carbonyl oxygen. An alternative mechanism in which the mobile proton is transferred to the amide nitrogen has also been proposed. Of the two proposed mechanisms, protonation at the carbonyl oxygen is more logical because the amide nitrogen is less basic than the carbonyl oxygen of an amide bond. With the protonation of the carbonyl oxygen, a nearby carbonyl can act as a nucleophile to

attack the electropositive carbon of the protonated carbonyl forming the five-membered oxazolone intermediate. The cyclic intermediate rapidly progresses through an elimination step to form the b_3^+ ion.

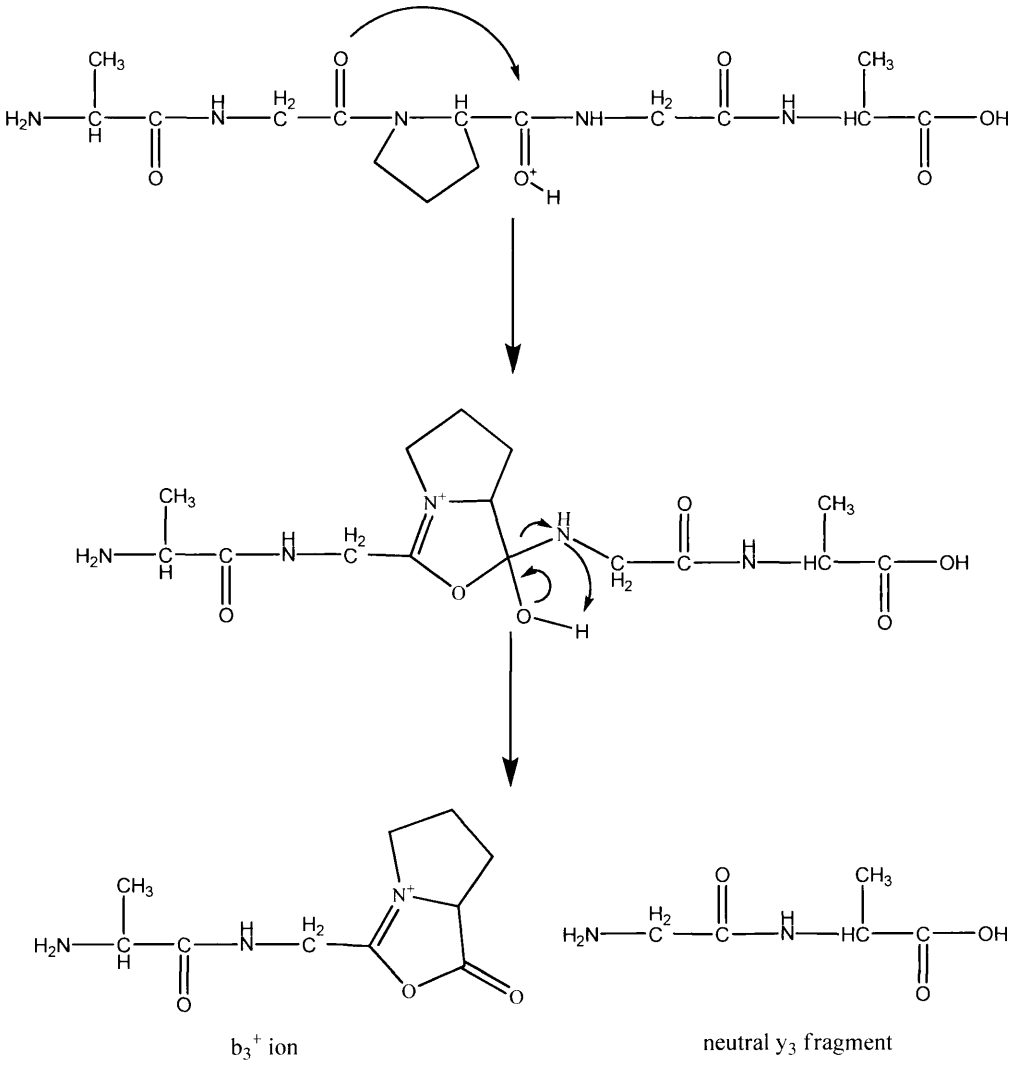


Figure 5.4 – Proposed Mechanism for b_3^+ Ion Formation[56]

In the formation of the y_3^+ ion, the amide oxygen N-terminal to proline is protonated by the mobile hydrogen, triggering a nucleophilic attack on the carbon

center of the protonated carbonyl by the amide oxygen of the second residue N-terminal to proline. The resulting five-membered ring oxazolone intermediate is then free to proceed through two divergent rearrangement pathways to ultimately form the y_3^+ ion. These pathways are shown in Figures 5.5 and 5.6. The proposed mechanism by Paizs and Suhai[56] provides a foundation for further investigations into the "proline effect".

In order to facilitate the further understanding of the governing factors of the "proline effect", a joint study with Wysocki and coworkers was undertaken. Recent studies by Poutsma and coworkers have investigated the effect of ring size on the proton affinity of proline analogs. Proton affinities for the amino acid homologs of heterocyclic nitrogen bases, azetidine-2-carboxylic acid (Aze), proline (Pro), and pipecolic acid (Pip) have been determined. A slight monotonic increase in the proton affinities (933.0, 941.0, 944.0 kJ/mol) was seen.[14] It was found that peptides containing Aze or Pro residues fragment to form y_3^+ ions whereas those containing Pip or the acylic N-methylalanine residue fragment to produce b_3^+ ions.[56] Production of b_3^+ or y_3^+ ions necessarily involves transfer of the mobile proton to different carbonyl oxygens. To see if the N-terminal carbonyl oxygen in Pro- or Aze-containing peptides were significantly more basic than those in Pip- or NMA-containing peptides, we calculated the absolute PAs of the peptide models $CH_3C(=O)$Aze and analogs.

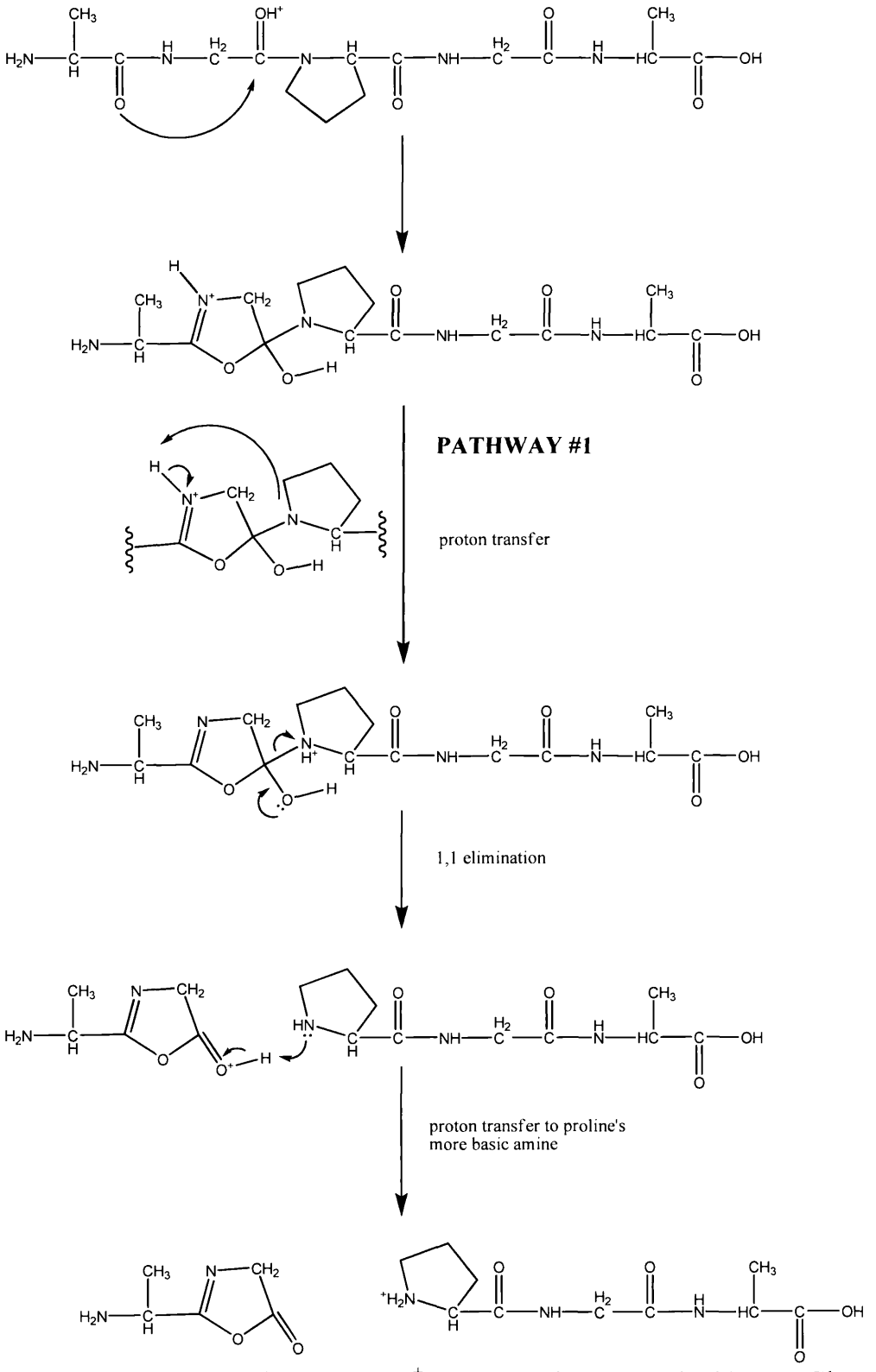


Figure 5.5 – Proposed Mechanism for y_3^+ Ion Formation from Ala-Gly-Pro-Gly-Ala Via Pathway #1[56]

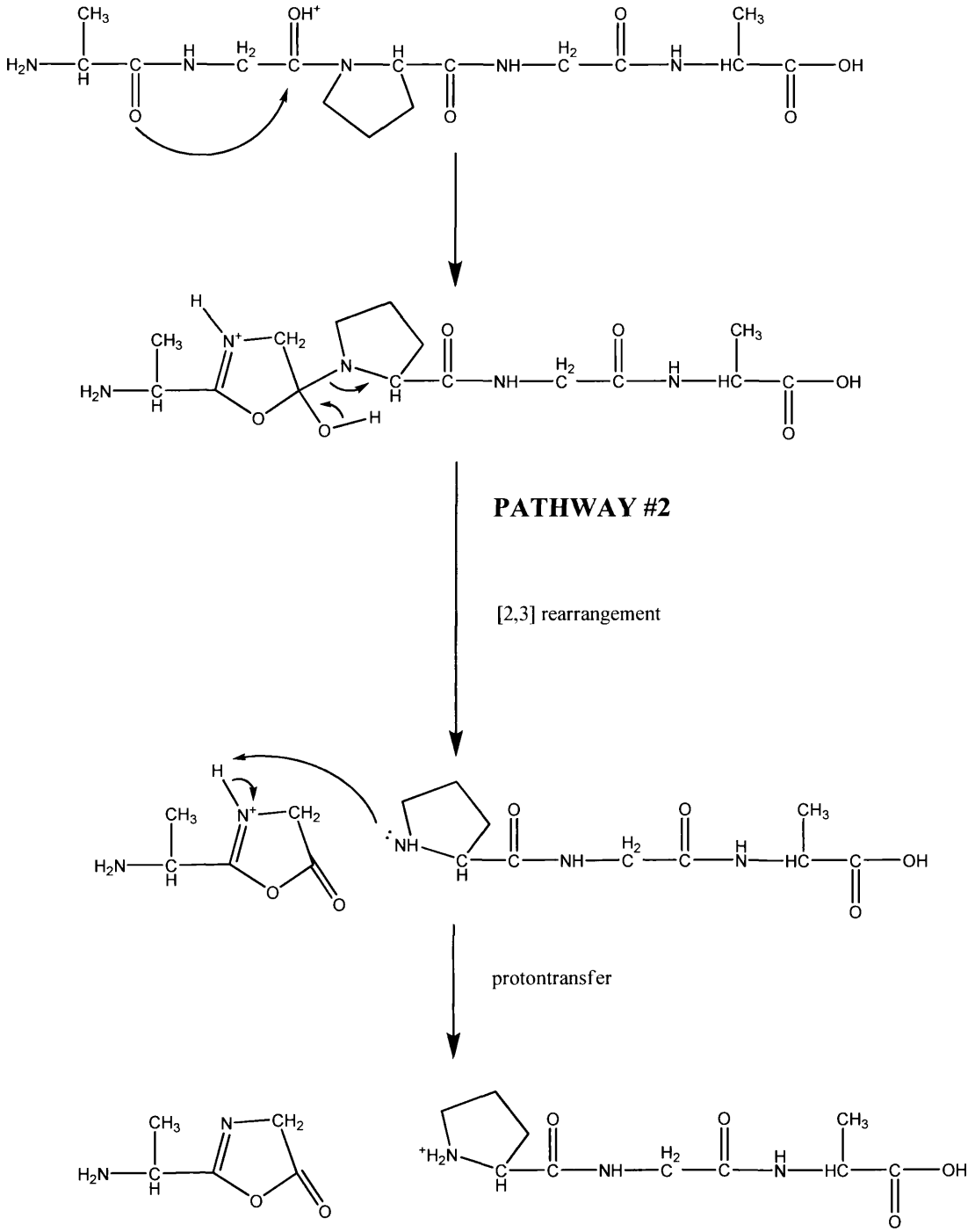
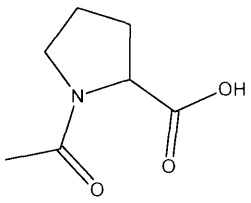


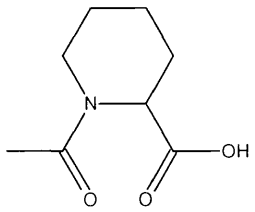
Figure 5.6 – Proposed Mechanism for y_3^+ Ion Formation from Ala-Gly-Pro-Gly-Ala Via Pathway #2[56]

5.3 Computational Calculations of Acetylated Proline Analogs

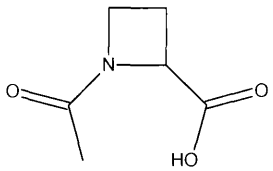
The mobile proton model of peptide fragmentation states that upon ionization, the proton(s) added to a peptide will migrate to various protonation sites prior to fragmentation. Upon collision-induced dissociation, fragmentation will occur at these protonation sites. Molecular orbital calculations on various protonated forms have shown that protonation on the amide nitrogen leads to a considerable weakening of the amide bond.[48] In an effort to understand the governing factors of the "proline effect", theoretical calculations were carried out with the purpose of obtaining absolute proton affinities for acetylproline, acetylpipecolic acid, acetylazetidine-2-carboxylic acid, and acetylNMA (Figure5.7).



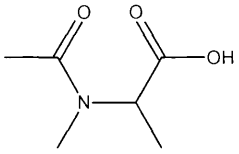
Acetylproline



Acetylpipecolic acid



Acetylazetidine-2-carboxylic acid



AcetylNMA

Figure 5.7 – Neutral Structures of Acetylated Proline Analogs

The theoretical analysis of each acetylated proline analog followed the procedure as laid out in Section 3.2. The process was repeated for the above molecules protonated at the amide nitrogen and amide carbonyl oxygen. Relative proton affinities for the various protonation sites were obtained directly from 298K enthalpies for the neutral and protonated molecules according to equation (18), where A and AH^+ are the neutral and protonated molecule and $H_{298}(H^+)$ was taken as $\frac{5}{2}RT$ = 12.3 kJ/mol.

$$PA = H_{298}(A) + H_{298}(H^+) - H_{298}(AH^+) \quad (18)$$

Computational studies of acetylproline, acetylpipecolic acid, acetylazetidine-2-carboxylic acid, and acetylNMA were carried out using the Gaussian98® and Gaussian98W® suites of programs.[33] The calculations were also repeated for the above molecules protonated at the amide carbonyl oxygen and amide nitrogen.

5.4 Results and Discussion

Proline is unique among amino acids in that its cyclic structure produces unique conformational properties on the peptide backbone. The ring has been shown to provide the amino acid with a natural turn believed to create the folding organization needed for higher protein formation. The phenomenon known as the "proline effect" is a commonly observed fragmentation preference for cleavage of the amino-terminal to the proline residues within a peptide. Also, it has been noted that the amino-terminal proline fragment takes the charge more often and is more abundant in mass spectra.

Absolute proton affinities were once again calculated, using an isodesmic approach with N,N dimethylacetamide (PA = 908 kJ/mol)[34] serving as the isodesmic reference compound according to equation (19).

$$A + C_4H_{10}NO^+ \longrightarrow AH^+ + C_4H_9NO \qquad (19)$$

Table 5.1 displays these calculated proton affinities and Figures 5.7-5.10 illustrate the lowest energy conformer for neutral and protonated acetylproline, acetylpipecolic acid, acetylazetidine-2-carboxylic acid, and acetylNMA. Theoretical thermochemical values for the lowest energy conformers of neutral and protonated species investigated in this work are listed in Table 5.2.

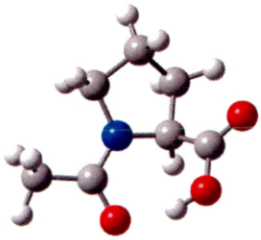
Table 5.1 – Absolute Proton Affinities of Proline Analogs

Structure	**Hybrid DFT Proton Affinity** (kJ/mol)
AcetylNMA (C=O)	905.3 ± 10.0
AcetylNMA (NH)	866.5 ± 10.0
Acetylazetidine-2-carboxylic Acid (C=O)	916.1 ± 10.0
Acetylazetidine-2-carboxylic Acid (NH)	855.9 ± 10.0
Acetylproline (C=O)	914.5 ± 10.0
Acetylproline (NH)	848.8 ± 10.0
Acetylpipecolic Acid (C=O)	913.2 ± 10.0
Acetylpipecolic Acid (NH)	864.8 ± 10.0

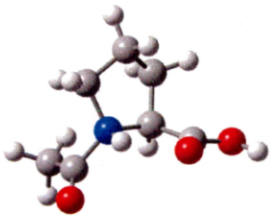
Table 5.2 – Total Electronic Energies, Thermal Corrections and 298K Enthalpies (Hartrees) for Selected Serine Analogs

Molecule	total electronic energy[a]	thermal correction[b]	ΔH_{298}
AcetylNMA (C=O)	-516.2297561	0.199830	-516.0299261
AcetylNMA (NH)	-516.2150766	0.199959	-516.0151176
Acetylazetidine-2-carboxylic Acid (C=O)	-515.0013348	0.176519	-514.8248158
Acetylazetidine-2-carboxylic Acid (NH)	-514.9784677	0.176603	-514.8018647
Acetylproline (C=O)	-554.3542058	0.207397	-554.1468088
Acetylproline (NH)	-554.3293238	0.207651	-554.1217898
Acetylpipecolic Acid (C=O)	--593.6765396	0.238221	-593.4383186
Acetylpipecolic Acid (NH)	-593.6579770	0.23807	-593.4199070

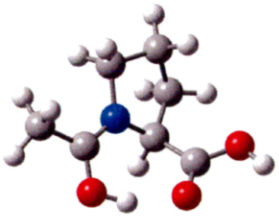
[a] B3LYP/6-311++G**//B3LYP/6-31+G* [b] ZPE + $\int c_p dT$, from unscaled harmonic frequencies obtained at B3LYP/6-31+G*



Neutral Acetylproline

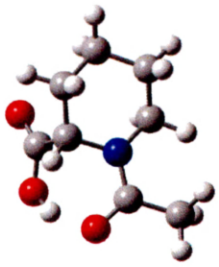


Acetylproline Protonated on Amide Nitrogen

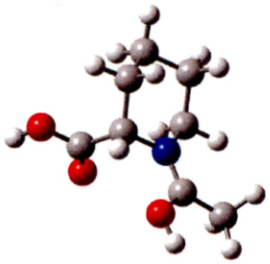


Acetylproline Protonated on Amide Carbonyl Oxygen

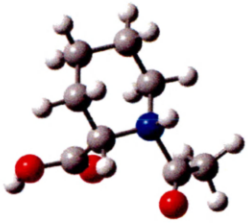
Figure 5.8 – Geometrically Optimized Neutral and Protonated Acetylproline Structures



Neutral
Acetylpipecolic Acid

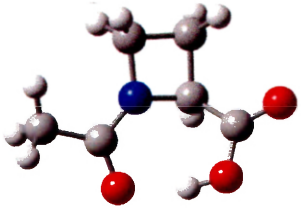


Acetylpipecolic Acid
Protonated on
Carbonyl Oxygen

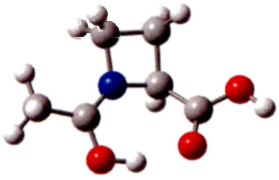


Acetylpipecolic Acid
Protonated on Amide
Nitrogen

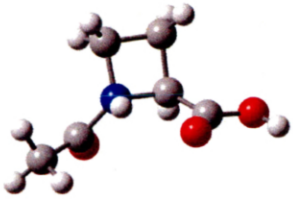
Figure 5.9 – Geometrically Optimized Neutral and Protonated Acetylpipecolic Acid Structures



Neutral
Acetylazetidine-2-
carboxylic Acid

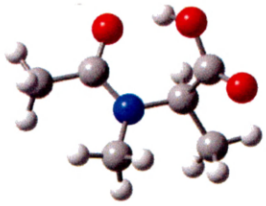


Acetylazetidine-2-
carboxylic Acid
Protonated on Amide
Carbonyl Oxygen

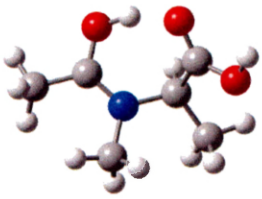


Acetylazetidine-2-
carboxylic Acid
Protonated on Amide
Nitrogen

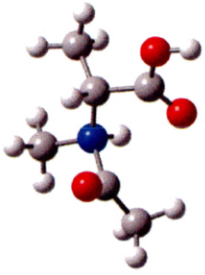
Figure 5.10 – Geometrically Optimized Neutral and Protonated Acetylazetidine-2-carboxylic Acid Structures



Neutral AcetylNMA



AcetylNMA
Protonated on Amide
Carbonyl Oxygen



AcetylNMA
Protonated on Amide
Nitrogen

Figure 5.10 – Geometrically Optimized Neutral and Protonated AcetylNMA Structures

In the three amino acid homologs of heterocyclic nitrogen bases, as well as acetylNMA, the amide carbonyl oxygen has a larger proton affinity than that of the amide nitrogen. Therefore, the increase in ring size shows no signs of influencing the preferential protonation site. These findings are in agreement with the study of Paizs and Suhai[57] in which a mechanism (Figure 5.4) for b_3^+ and y_3^+ ion formation from Ala-Gly-Pro-Gly-Ala has been proposed. In this mechanism, an initial transfer of the mobile proton to the carbonyl oxygen triggers a nucleophilic attack on the carbon of

the protonated carbonyl. A five-membered oxazolone intermediate is produced and rapidly progressed through an elimination step to form the b_3^+ ion.

As can be observed in Table 5.1, the proton affinities of the carbonyl oxygen atoms in the proline homolog residues were within the experimental uncertainties of ± 10 kJ/mol of each other. From this, the assumption can be made that the proton affinities are not the driving force behind the selective fragmentation pathways. The current study suggests that it is the rigidity of the ring structure that promotes the b_3^+ or y_3^+ ion formation. To further validate this claim, future studies on N-acetylproline-N-methyl amide (Figure 5.11) and its analogs are a necessity.

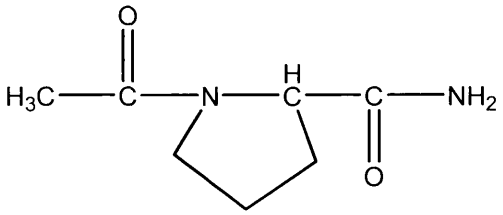


Figure 5.11 – Neutral Structure of N-acetylproline-N-methyl amide

This future study would allow for the determination of the proton affinities of the various carbonyl groups and ultimately clarify the role, if any, of the ring size and stability in the fragmentations mechanisms.

References

[1]Berg, J.M.; Tymoczko, J.L.; Stryer, L. *Biochemistry*; Freeman: New York, **2007**
[2] Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. *The FASEB Journal*, **1996**, *10*, 75.
[3] Griffiths, J. *Anal. Chem.* **2008**, 80, 5678.
[4]Aston, F.W. *Phil. Mag.* **1919**, 38, 707.
[5]Aston, F.W. *Phil. Mag.* **1920**, 39, 611.
[6]Dempster, A.J. *Phys. Rev.* **1918**, 11, 316.
[7]Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press: San Diego, **2003**.
[8]Paul, W. *Rev. Mod. Phys.* **1990**, 62, 531.
[9]Hoffmann, E.; Stroobant,V. *Mass Spectrometry*; Wiley: England, **2007**
[10]Sternet,J.L.; Johnston, M.V.; Nicol, G.R.; Ridge, D.P. *J. Am. Soc. Mass Spectrom.* **1999**, 10, 483.
[11]Harrison, A.G. *Mass Spectrom. Rev.* **1997**, 16, 201.
[12]Rosenthal, G.A. *Amino Acids.* **2001**, 21, 319.
[13]Bell, E. A. *J. Agric. Food. Chem.* **2003**, 51, 2854.
[14]Kuntz, A.F.; Boynton, A.W.; David, G.A.; Colyer, K.E.; Poutsma, J.C. *J. Am. Soc. Mass Spectrom.* **2002**, 13, 72.
[15]Schroeder, O.E.; Andriole, E.J.; Carver, K.L.; Colyer, K.E.; Poutsma, J.C. *J. Phys. Chem.* **2004**, 108, 326.
[16]James, P. *Q. Rev. Biophys.* **1997**, 30, 279.
[17]Cooks, R. G.; Denault, J. W.; Williams, T. I. *Int. J. Mass Spectrom.* **2001**, 210/211, 133.
[18]Robinson, P. J.; Holbrook, K. A. *Unimolecular Reactions*; Wiley-Interscience: New York, **1972**
[19]Cooks, R.; Zheng, X. *J. Phys. Chem. A* **2002**, 106, 9939.
[20]Cooks, R. G.; Wong, P. S. H. *Acc. Chem. Res.* **1998**, 31, 379.
[21]Cheng, X. H.; Wu, Z. C.; Fenselau, C. J. *Am. Chem. Soc.* **1993**, 115, 4844.
[22]Cerda, B. A.; Wesdemiotis, C. *J. Am. Chem. Soc.* **1998**, 120, 2437.
[23]Denault, J.W. *Temperature corrected kinetic method for the estimation of proton affinities, entropies of protonation and gas phase basicities*; Purdue University, **2000**.
[24]Armentrout, P. B. *J. Am. Soc. Mass. Spectrom.* **2000**, 11, 371.
[25]Ervin, K. M.; Armentrout, P. B. *J. Mass. Spectrom.* **2004**, 39, 1004.
[26]Jones, M. C.; Bernier, M.; Carson, E.; Colyer, K. E.; Metz, R.; Pawlow, A.; Wischow, E. D.; Webb, I.; Andriole, E. J.; Poutsma, J. C. *Int. J. Mass Spectrom.* **2007**, 267, 54.
[27]Andriole, E. J.; Colyer, K. E.; Cornell, E.; Poutsma, J. C. *J. Phys. Chem. A* **2006**, 110, 11501.
[28]Levine, I. N. *Physical Chemistry*; McGraw-Hill: New Delhi, **2002**.
[29]Cramer, C. J. *Essentials of Computational Chemistry*; Wiley: England, **2002**.
[30]Young, D. *Computational Chemistry*; Wiley-Interscience: New York, **2001**.

[31]Becke, A. D. *J. Chem. Phys.*, **1993**, *98,* 5648.
[32]Lee, C.; Yang, W.; Parr, R.G. *Phys. Review*, **1988**, *37,* 785
[33]Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, N.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, version A.9; Gaussian, Inc: Pittsburgh, PA, **1998**
[34]Hunter, E.P.; Lias, S.G. *J. Phys. Chem. Ref. Data*, **1998**, 27, 413.
[35]Tian, Z.; Pawlow, A.; Poutsma, J.C.; Kass, S.R. *J. Am. Chem. Soc.* **2007**, 129, 5403.
[36]Gronert, S.; Simpson, D.C.; Conner, K.M. *J. Am. Soc. Mass Spectrom.* **2009**, 20, 2116.
[37]O'Hair, R.J.; Bowie, J.H.; Gronert, S. *Int. J. Mass Spectrom. Ion. Proc.* **1992**, 117,23.
[38]Bleiholder, C.; Suhai, S.; Paizs, B. *J. Am. Soc. Mass Spectrom.* **2006**, 17, 1275.30
[39]Nesvizhskii, A. I.; Aebersold, R. *Mol. Cell. Proteomics*. **2005**, *4*, 1419.
[40]Harrison, A. G.; Young, A. B. *J. Mass Spectrom*. **2005**, 40, 1173.
[41]Wysocki, V.H.; Tsaprailis, G.; Smith, L.L; Breci, L. A. *J. Mass Spectrom.* **2000**, 35, 1399.
[42]Rožman, M. *J. Am. Soc. Mass Spectrom.* **2007**, 18, 121.
[43]Roepstorff P.; Kohlman, J. *Biomed Mass Specrom*, **1984**, *11*, 601.
[44]Biemann K. *Annu Rev Biochem*, **1992**, *61*, 977.
[45]Leymarie, N.; Berg, A.E.; McComb, M.E.; O'Connor, P.B.; Grogan, J.; Oppenheimer, F.G.; Costello, C.E. *Anal. Chem.* **2002**, 74, 4124.
[46]Schwartz, B. L.; Bursey, M. M. *Biol. Mass Spectrom*. **1992**, 21, 92.
[47]Vaisar, T.; Urban, J. *J. Mass. Spectrom*. **1996**, 31, 1885.
[48]McCormack, A.L.; Somogyi, Á.; Dongré, A.R.; Wysocki, V.H. *Anal. Chim. Acta*. **1993**, 65, 2859.
[49]Dongré, A.R.; Jones, J.L.; Somogyi, Á.; Wysocki, V.H. *J. Am. Chem. Soc*. **1996**, 118, 8365.
[50]Vaisar, T.; Urban, J. *J. Mass Spectrom*. **1998**, 33, 505.
[51]Dongré, A.R.; Jones, J.L.; Somogyi, Á.; Wysocki, V.H. *J. Am. Chem. Soc.* **1996**, 118, 8365.
[52]Gu, C.G.; Somogyi Á.; Wysocki, V.H.; Medzihradszky, K.F. *Anal. Chim. Acta* **1999**, 397, 247.
[53]McCormack, A.L.; Somogyi, Á.; Dongré, A.R.; Wysocki, V.H. *Anal. Chem.* **1993**, 65, 2859 .

[54]Nair, H,; Wysocki, V.H. *Int. J. Mass Spectrom.* **1998**; 174, 95.
[55]Jones, J.L.; Dongré, A.R.; Somogyi, Á.; Wysocki, V.H. *J. Am. Chem.* Soc. **1994**, 116, 8368.
[56]Raulfs, M.D.M. *Investigations of the Mechanism of the "Proline Effect";* College of William and Mary, **2007**.
[57]Paizs, b.; Suhai, S. *Mass Spectrom. Rev.* **2005**, 24, 508.

Appendices

Appendix A – Supporting Documents for Homocysteine

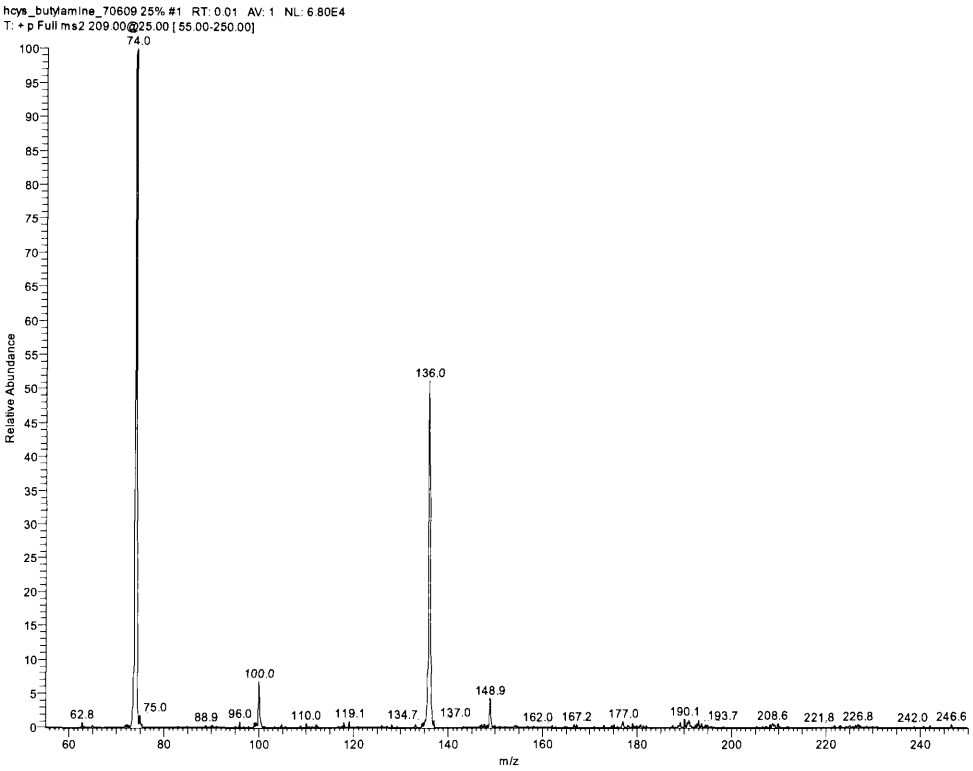


Figure A – Fragmentation Spectrum of Homocysteine-Butylamine

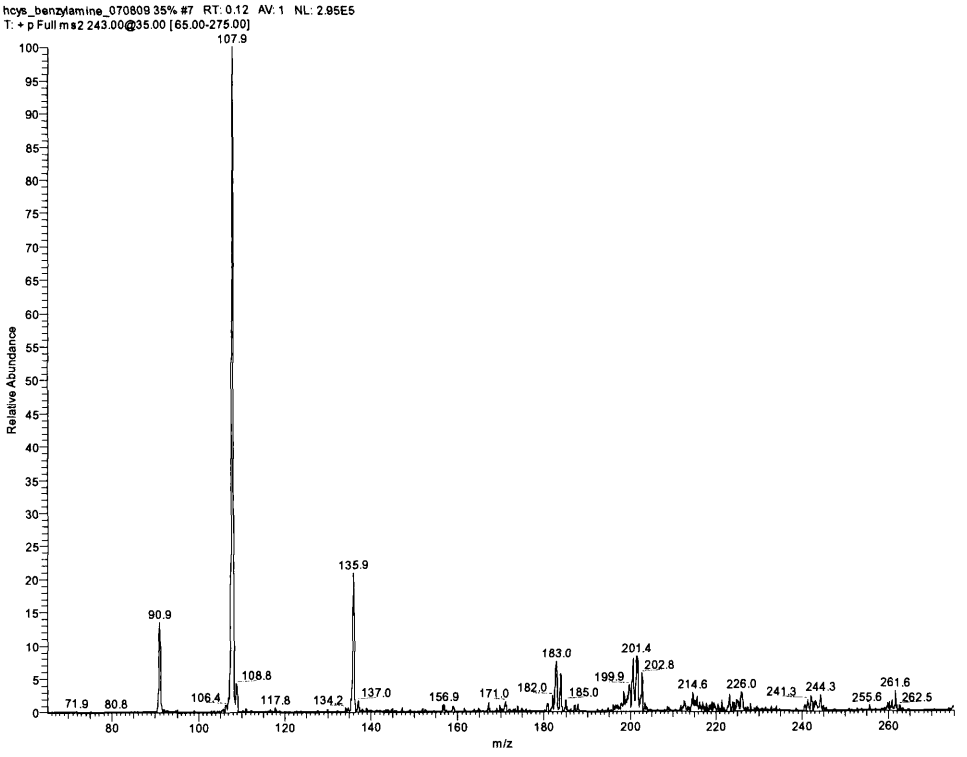


Figure B – Fragmentation Spectrum of Homocysteine-Benzylamine

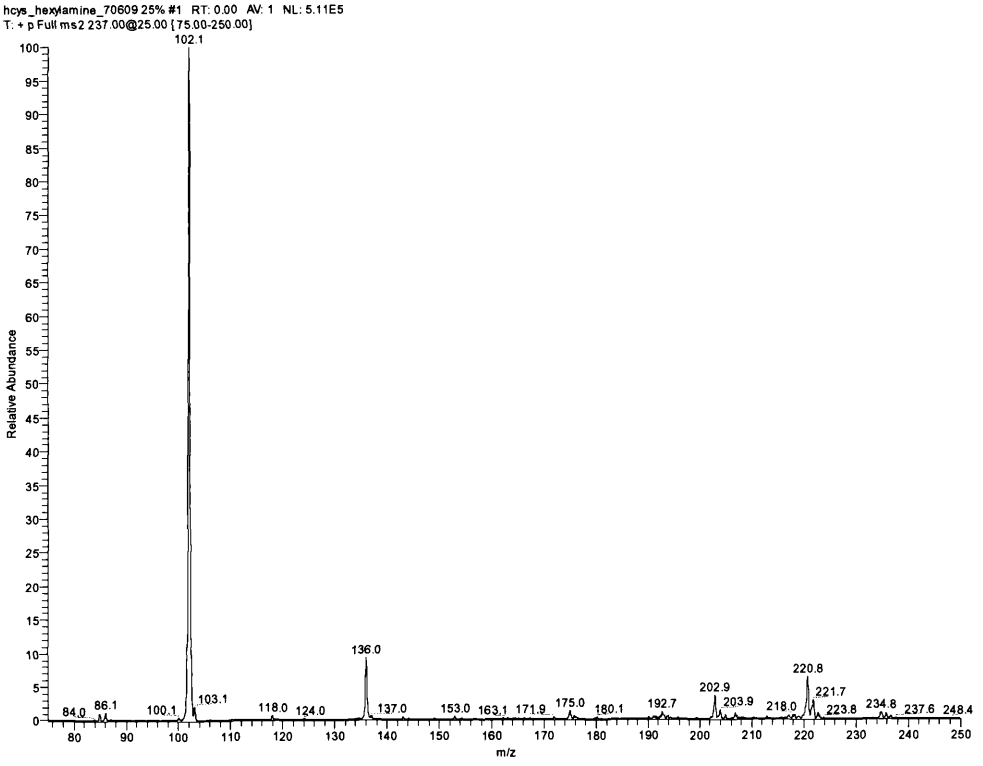


Figure C – Fragmentation Spectrum of Homocysteine-Hexylamine

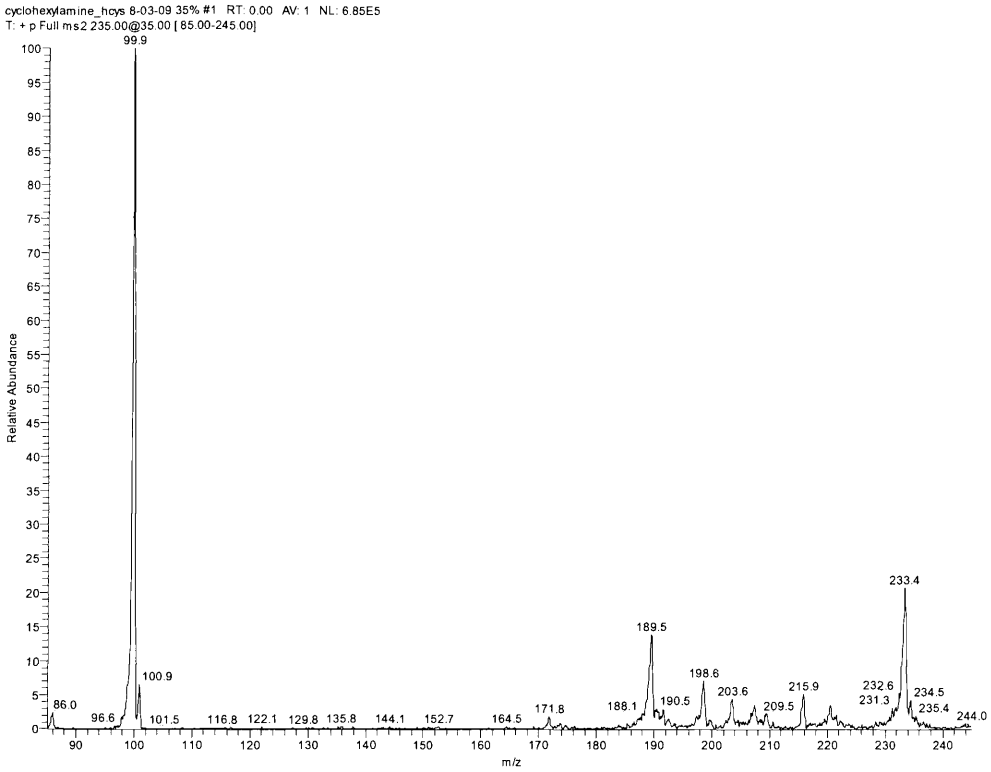


Figure D – Fragmentation Spectrum of Homocysteine-Cyclohexylamine

Appendix B – Supporting Documents for Homoserine

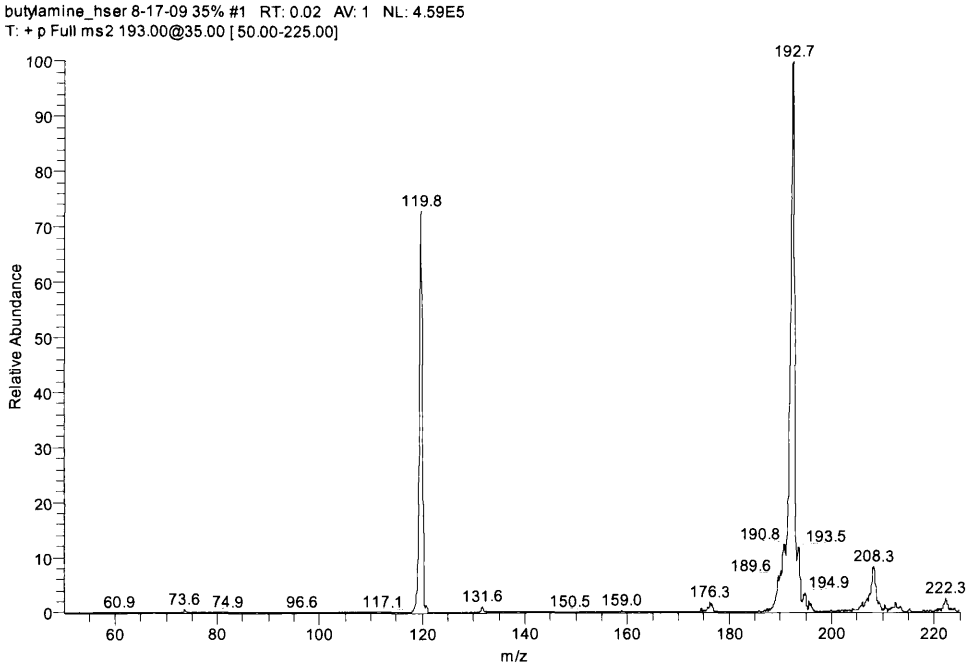


Figure E – Fragmentation Spectrum of Homoserine-Butylamine

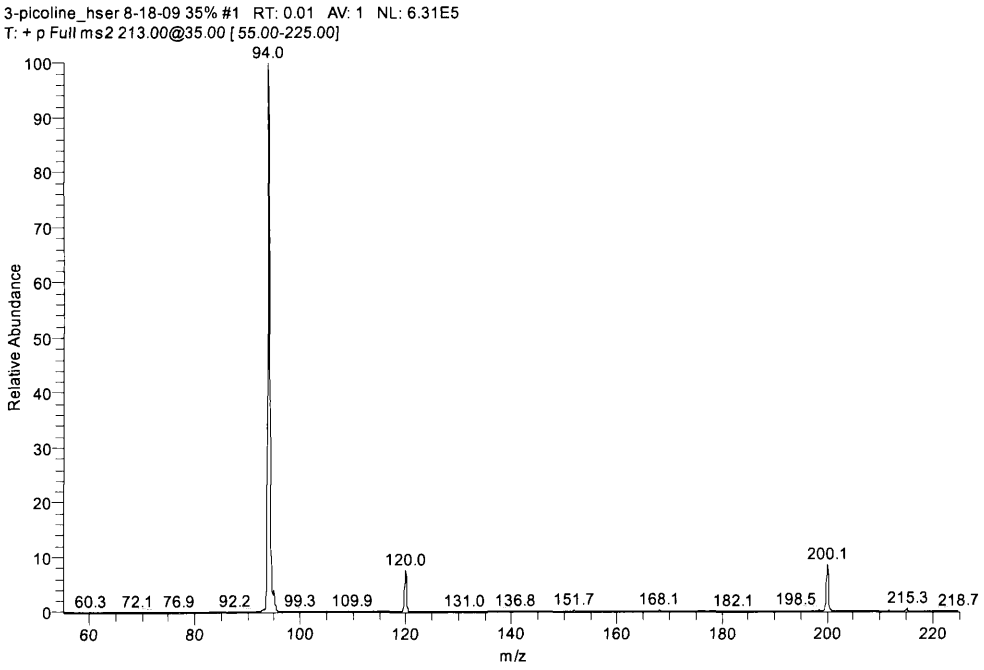


Figure F – Fragmentation Spectrum of Homoserine-3-Picoline

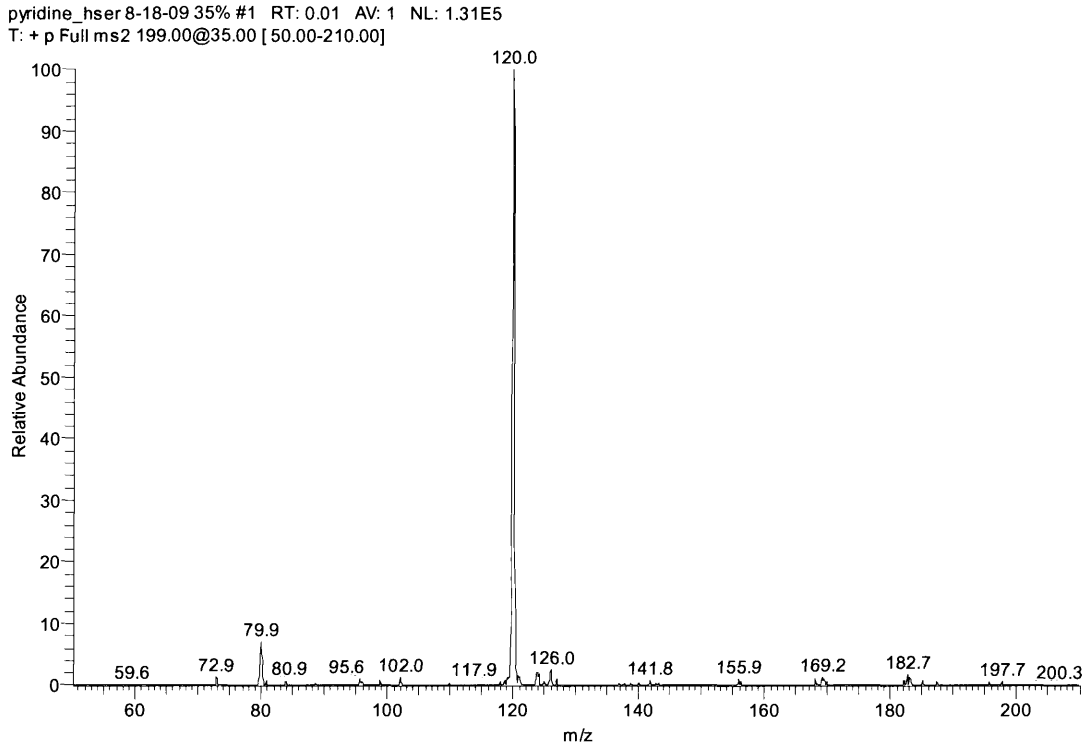


Figure G – Fragmentation Spectrum of Homoserine-Pyridine

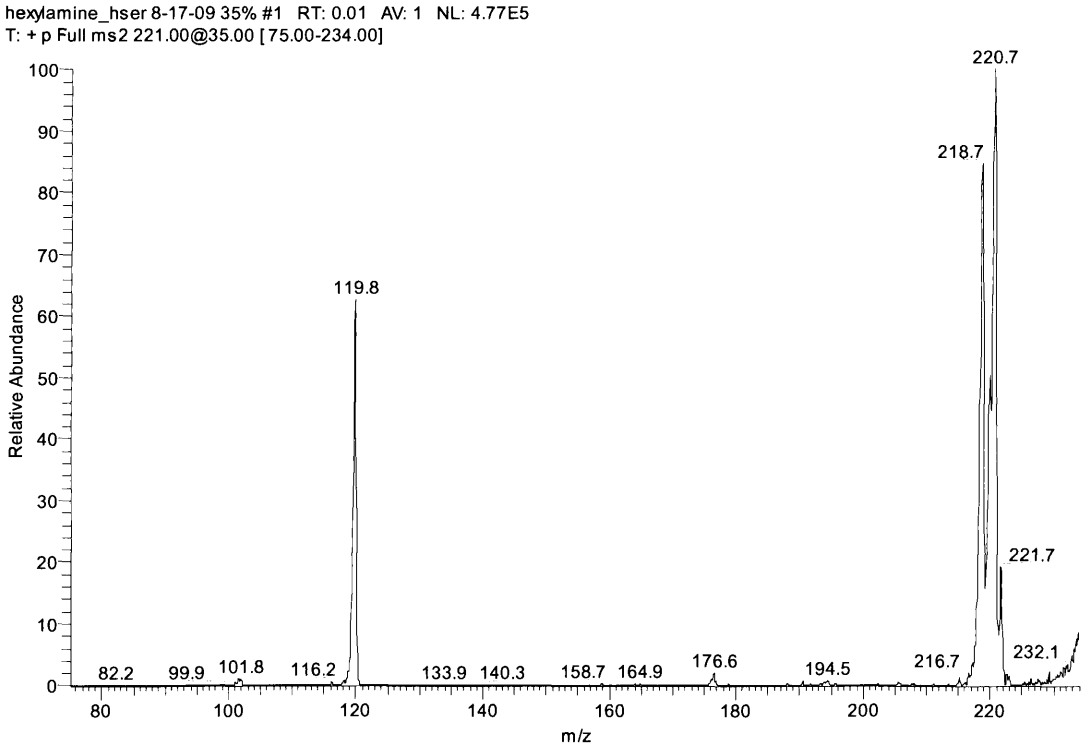


Figure H – Fragmentation Spectrum of Homoserine-Hexylamine